Quantitative and Population Genetic Analyses of Domesticated and Wild Sheep Populations

Allan F. McRae

Thesis submitted for the degree
of
Doctor of Philosophy

University of Edinburgh

2005

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own.

The work has not been submitted for any other degree or professional qualification except as specified.

Allan McRae

Acknowledgements

This thesis would not have been accomplished without the help and encouragement of a large number of people to whom I am very grateful.

Firstly, I would like to thank my primary supervisor Peter Visscher for his support over the last three years. His constructive criticism on many manuscripts and thesis chapters has greatly improved the quality of the research presented here. Also, thanks to Josephine Pemberton for her supervision of the work involving the Soay sheep as well as providing me the opportunity to visit St. Kilda.

A large number of people have been involved in the generation of the various data sets analysed in this thesis. Steve Bishop and Grant Walling organized the collection and genotyping of samples from the commercial sheep populations. Dario Beraldi provided genotypes for the Soay sheep pedigree and also some much needed direction in my adventures into the lab. The modelling of the Soay sheep population was based on data collected by numerous researchers and volunteers over many years.

My studies were made easier due to the support of family and friends. Thanks go to my parents who have supported me over my entire career as a student. I have had the pleasure of sharing an office with a number of people during my PhD, all of whom have provided many interesting discussions, sometimes work related, over the vast number of required tea breaks.

I would especially like to thank my wife Christine to whom I dedicate this thesis. Her constant encouragement over the past three years has made this possible. I am assured that this support was especially arduous over the last few months when I put fingers to keyboard into the small hours of the night.

This work was supported by the Association of Commonwealth Universities and the British Council through Commonwealth Scholarship number NZCR-2002-164.

Publications

The following publications have resulted as a direct outcome of the research described in this thesis:

- McRae, A. F., Bishop, S. C., Walling, G. A., Wilson, A. D. and Visscher, P. M. (2005). Mapping of multiple quantitative trait loci for growth and carcass traits in a complex commercial sheep pedigree. *Animal Science* **80:** 135-142. [Chapter 2]

- McRae, A. F. and Beraldi, D. (2006). Examination of a region showing linkage map discrepancies across sheep breeds. *Mammalian Genome* (in press). [Chapter 3]

- McRae, A. F., Pemberton, J. M. and Visscher, P. M. (2005). Modeling linkage disequilibrium in natural populations: The example of the Soay sheep population of St. Kilda, Scotland. *Genetics* **171:** 251-258. [Chapters 5 and 6]

The following publications have resulted from research associated with this thesis:

- Beraldi, D., McRae, A. F., Gratten, J., Slate, J., Visscher, P. M. and Pemberton, J. M. Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population. (submitted).

- Gratten, J., Beraldi, D., Lowder, B., McRae, A. F., Visscher, P. M., Pemberton, J. M. and Slate, J. The genetic basis of coat colour variation in a population of feral sheep. (submitted).

Abstract

Much of the phenotypic variation between individuals of a species is in the form of quantitative traits. To gain an understanding of the nature of this variation requires the dissection of their underlying genetic architecture and the examination of how environmental factors influence its expression. Livestock populations provide an interesting system for the study of the genetics of quantitative traits. Differences between breeds of a species are readily observed and a substantial amount of variation exists between and within populations of a particular breed. The existence of feral populations provides further opportunities as these allow the examination of complex evolutionary traits such as fitness. The combined use of feral and domestic populations also allows the investigation of the effects of livestock management on the genetic architecture and expression of quantitative traits.

In Chapter 2, a directed linkage scan for loci involved in body weight and carcass composition traits is performed in a commercial Charollais sheep population. Five chromosomes were investigated based on prior evidence for major genes affecting the studied traits in other breeds and species. A maximum likelihood variance component analysis using identity-by-descent values estimated by Markov chain Monte Carlo methods was perform on a complex pedigree containing a total of 570 sheep. Of the total of nine QTL detected, the estimated position of only one overlapped with the regions showing major genes that were used in chromosome selection.

During the analysis of the Charollais sheep population, a region of the genome showing a significant deviation from the published sheep linkage map was detected. This region is examined in more detail in Chapter 3, with the addition of further microsatellite markers as well as the investigation of this region in two further sheep breeds. With the inclusion of the published linkage map, this demonstrated a total of three linkage maps across four populations. Such heterogeneity in linkage maps across sheep breeds has important consequences for the design and analysis fine-mapping studies.

The significance of a QTL linkage peak is not readily evaluated with general pedigrees. The extension of permutation methodology that is commonly used with structured pedigrees to more general pedigrees is investigated in Chapter 4. Estimation of significance thresholds is achieved through the permutation of

independent groups within the pedigree that are formed subject to constraints on relationships to other pedigree members. The estimated thresholds showed substantial bias in cases where a small portion of the pedigree was permutable and in the presence of QTL with large effects.

Chapter 5 examines the population dynamics of a well studied wild Soay sheep population. A unified statistical framework is developed for all major aspects of the life cycle of the sheep. This forms the basis of a simulation model of the population that is used to predict the amount of linkage disequilibrium in the population (Chapter 6) and the effective population size of the population (Chapter 7). The examination of the linkage disequilibrium structure in a population is an important step in the design of studies with the aim of fine-mapping quantitative trait loci. The simulated population showed significant decline of linkage disequilibrium with genetic distance and low levels of background linkage disequilibrium, indicating that the Soay sheep population is a viable resource for linkage disequilibrium fine mapping. Through the use of the simulation model, the effective population size of the Soay sheep population was estimated to be approximately 0.17 of its census population size. This is approximately half the value obtained with the use of a general predictive equation.

Table of Contents

1 Quantitative Genetics in Sheep

1.1 Introduction

The major goal of quantitative genetics is to understand the nature of the variation in quantitative traits (Lynch and Walsh, 1998). Many of the differences between individuals of a species are in the form of quantitative traits. The expression of these traits involves a complex interaction between an individual's genotype and its environment. Thus, understanding the nature of quantitative traits beings at the level of the proportion of the variance in trait values of a population that is due to either genetic of environmental differences. If a genetic component is detectable, the genetic architecture can then be further dissected. The genetic architecture refers to the number of genes that influence a particular phenotype, the effect size of different allele at these genes and how these genes interact with each other and the environment to ultimately produce the trait of interest (Zeng *et al.*, 1999; Mackay, 2001). Knowledge of the genetic architecture of a quantitative trait can be used in answering general questions about the nature of quantitative traits, such as the mechanisms for the maintenance of genetic variation in natural populations and the responses of quantitative traits to selection (Barton and Keightly, 2002).

This thesis focuses on quantitative genetics in sheep. Sheep were among the earliest of livestock species to be domesticated, with domestication occurring approximately 10,000 years ago along with the goat (Franklin, 1997; Simm, 1998). The early domestication of the sheep may be attributed to their hardiness and versatility. Sheep are able to utilize low quality pastures and survive in adverse climatic conditions. There hardiness is demonstrated by the wide variety of habitats in which they are found, ranging from cold mountainous regions to hot deserts.

The versatility of sheep continues their commercial importance today. Sheep provide two sources of food in their meat and milk. Lamb and mutton collectively represent approximately 4% of meat eaten worldwide. Their milk is utilized for the production of cheese and yoghurt, especially in Mediterranean countries. Sheep also provide a valuable source of textile fibre, with wool being the primary source of fibre of animal origin. In poorer countries, sheep also provide pelts, fertilizer and fuel.

This chapter provides an overview of the state of quantitative genetic studies in sheep. Firstly, the genome structure of sheep is explored with emphasis on comparisons to the genomes of other species which provide a valuable resource for quantitative genetic studies. This is followed by a brief overview of methodology used in the mapping of quantitative traits loci (QTL) and a summary of QTL mapping experiments that have been performed in sheep. Finally, an overview of the remainder of this thesis is presented.

1.2 Phylogenetics and Comparative Genomics

Scientific Classification: The scientific classification of the sheep is given in Table 1.1. The order Artiodactyla contains ten extant families and approximately 200 species (Franklin, 1997) making it one of the most successful mammalian orders. The suborder classification depends on what level the tragulinids (mouse deer) split from the remaining (Pecoran) ruminants. Pecora could equally be considered an infraorder of Ruminatia. The family Bovidae contains important commercial species other than sheep, including cattle and goats, as well as many species native to Africa such as gazelles and wildebeest. Other families of commercial importance in the order Artiodactyla are Suidae (in suborder Suiformes) containing pigs, Camelidae (in suborder Tylopoda) containing llamas

Table 1.1: Scientific classification of the domestic sheep. Adapted from Frankham, 1997.

Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Artiodactyla
Suborder:	Pecora (Ruminantia)
Family:	Bovidae
Subfamily:	Caprinae
Genus:	*Ovis*
Species:	*aries*

and Cervidae (in suborder Pecora) containing deer.

The classification within genus *Ovis* is somewhat controversial with the number of species ranging from four to eight (see Table 1.2). Much of this controversy extends from the fact that all the "species" can interbreed and thus could be considered races of the same species (Franklin, 1997). However, some support for the species level comes from the very apparent division between races from Eurasia (argali, urial and mouflon) and America (bighorn, thinhorn and snow sheep) and the varying chromosome number within these groupings. The origins of domestic sheep remain uncertain. Originally, the urial was considered a candidate ancestor due to its presence in the region where sheep were first domesticated. However, cytogenetic studies showed the urial had a diploid chromosome number of 58, greater than the 54 observed in domestic sheep. Further studies suggested the mouflon as a potential ancestor as it also has a diploid chromosome number of 54. This information was far from conclusive as it was found hybrid ewes from matings between argali and mouflon produced

Table 1.2: Division of the genus *Ovis*. Adapted from Ryder, 1984.

Species	Common Name
Ovis ammon	Argali
Ovis aries	Domestic sheep
Ovis canadensis	Bighorn
Ovis dalli	Thinhorn
Ovis musimon	European mouflon
Ovis nivicola	Snow sheep
Ovis orientalis	Asian mouflon
Ovis vignei	Urial

ovum with the lower haploid number of chromosomes, indicating strong selection toward reduced chromosome number. Recent molecular evidence suggests two clades within domestic sheep, one sharing a common ancestor with the European mouflon and one derived from another common ancestor for which the argali and urial are excluded (Hiendleder *et al.*, 1998; Hiendleder *et al.*, 2002). The two grouping of domestic sheep show a strong but incomplete correlation with modern fat- and thin-tailed domestic varieties.

As with other species, the domestication of sheep has resulted in a much greater phenotypic variety than observed in their wild counterparts (Simm, 1998). There are between 800 and 1000 breeds of domestic sheep (Loftus and Scherf, 1993; Mason, 1996) reflecting the versatility of sheep as a species. The classification of populations into breeds is typically based on uniformity of particular physical characteristics but other factors such as geographical location are also an influence.

Genome Structure: The domestic sheep genome consists of 26 pairs of autosomes

and two sex chromosomes, giving a diploid number of 54. Chromosomes 1, 2 and 3 are very large compared to other chromosomes having genetic lengths greater that 300cM, more than double the length of any other chromosome. These chromosomes are also unique in that they are all metacentric while the remainder of the autosomes are telocentric. The X chromosome is acrocentric and relatively large, having genetic length similar to that of the largest telocentric autosome. The Y chromosome is very small and metacentric. The sheep linkage map has undergone several revisions (Crawford *et al.*, 1995; Galloway *et al.*, 1996; de Gortari *et al.*, 1998; Maddox *et al.*, 2001) and now contains 1062 loci covering 3,400cM (sex-averaged) for the autosomes and 132cM (female) for the X chromosome.

Comparative Maps: The closest relative of sheep that has a well defined linkage map is the goat, *Capra hircus*, which is also classified in the subfamily Caprinae. The goat genome contains 29 autosomal pairs, three more than the sheep genome. The published goat linkage map is currently in its second generation (Vaiman *et al.*, 1996; Schibler *et al.*, 1998), and consists of 307 markers and covering approximately 88% of the genome. The chromosome coverage is still far from complete, with sixteen chromosomes containing greater than one linkage group or markers unlinked to the primary linkage group. Of the 307 markers in the goat linkage map, 218 (71%) are common to the sheep linkage map. The three large metacentric chromosomes in the sheep genome are equivalent to two goat chromosomes each, with sheep (OAR) chromosome 1 being equivalent to goat (CHI) chromosomes 1 and 3, OAR 2 being equivalent to CHI 2 and 8, and OAR 3 to CHI 5 and 11. A comparison of the sheep and goat linkage maps finds only four mapped loci occurring on goat chromosomes other than predicted (Maddox, 2004). There is also a large number of inversions between the chromosomes but some of these may be attributed to the lack of robustness

in the construction of the goat linkage map caused through the low numbers of markers (Maddox, 2004).

Within the family Bovidae, a further comparison can be made to the cattle (*Bos tauros*) genome. The cattle genome contains of 29 autosomal pairs, the same number as the goat. The cattle linkage map has undergone several revisions (Barendse *et al.*, 1994; Bishop *et al.*, 1994; Ma *et al.*, 1996; Barendse *et al.*, 1997; Kappes *et al.*, 1997; Ihara *et al.*, 2004) and is the most comprehensive linkage map for mammalian livestock. The latest revision contains 3960 markers, covering 3160cM in the (sex-averaged) autosomal linkage groups and the (female) X linkage group. Many links to the sheep linkage map are present with 572 markers found in both linkage maps. Similar to the goat genome, the three large chromosomes in the sheep genome are found to equivalent two two cattle chromosomes with OAR 1 being equivalent to cattle (BTA) chromosomes 1 and 3, OAR 2 to BTA 2 and 8, and OAR 3 to BTA 5 and 11. This indicates OAR 1 to 3 were formed by fusion events since the splitting of the sheep and goat lineages. A further chromosomal rearrangement between sheep and cattle is in the form of a Robertsonian translocation with OAR9 comprising of BTA14 and the telomeric end of BTA9. The remainder of BTA9 corresponds to OAR8. A comparison of the cattle and sheep genomes show less inversions than that of the sheep and goat genomes (Maddox, 2004), although, as discussed earlier, this is likely to be due to the lack of robustness in the goat genome construction.

Continuing further back into the evolutionary clades, the deer (subfamily Cervidae) is a member of the Pecora suborder. A hybrid pedigree of red deer (*Cervus elaphus*) and Père David's deer (*Elaphurus davidianus*) has been used in the construction of a deer linkage map (Slate *et al.*, 2002a). Markers used in the linkage map construction were chosen to allow the comparison to other ruminant

species and model organisms. This allowed the reconstruction of the karyotype of the common Pecoran ancestor of sheep, cattle and deer. In terms of major rearrangements, the lineage from this Pecoran ancestor the the Bovid ancestor required one fission. The linkage from the Pecoran ancestor to the deer requires 6 fissions, 1 fusion, 1 inversion and a possible translocation (Slate *et al.*, 2002a).

Comparisons to more distant relatives are hindered due to the low number of markers in the sheep linkage map that are conserved at across evolutionary time scales of this magnitude. Some progress has been made on the human-sheep comparative map using chromosomal painting (Burkin *et al.*, 1997; Iannuzzi *et al.*, 1999). These studies found a total of 48 human chromosome segments conserved in the sheep genome. Another approach to the construction of sheep-human comparative maps is to use the cattle map as a intermediate step. The cattle-human comparative map is much more defined due to the larger number of expressed sequence tags (ESTs) available for the cattle genome. Several cattle-human comparative maps have been published (Band *et al.*, 2000; Williams *et al.*, 2002; Larkin *et al.*, 2003; Everts-van der Wind *et al.*, 2004). The latest revision covers approximately 66% of the human genome and finds 195 segments (of more than two genes) conserved between the human and cattle genomes.

1.3 Methodology for Quantitative Trait Loci Mapping

Sheep show substantial phenotypic variation, both between and within breeds. Much of this variation is in the form of quantitative traits. To gain an understanding of the nature of this variation ultimately requires the dissection of the underlying genetic architecture and the examination of how environmental factors influence its expression. From a commercial standpoint, this knowledge is useful in that it would allow the marker assisted introgression of favourable alleles of a particular gene into a new population and the selection of favourable

genotypes within a population. The basic methodology for the localization and subsequent identification of genetic variants underlying quantitative traits as applicable to sheep populations is outlined below.

Linkage Mapping: Genetic linkage analysis is the primary method for the detection of quantitative trait loci (QTL). The basis premise is to identify regions of the genome whose segregation pattern shows a correlation with the phenotypic values of a trait of interest. The design of QTL mapping experiments in sheep is limited through their reasonably long gestation time and the small number of offspring afforded by each mating. A widely used design involves large half-sib families, created though the mating (often through artificial insemination) of a large number of dams by a single sire. Given the sire is segregating for a QTL, a single half-sib family is the most powerful for a given number of individuals (Weller *et al.*, 1990). However, a single sire is not likely to be segregating at all QTL and thus will not be informative at all loci. This results in a trade-off being made, with family sizes being reduced to allow several half-sib families to be created. The number of QTL segregating within each sire can be increased by using sires that result from matings between two breeds that differ greatly for the trait of interest or from "high" and "low" selection lines within a breed.

In its simplest form, the testing for linkage between a single marker and a QTL in a half-sib family can be achieved using a simple linear regression. Let the alleles of a marker that is heterozygous in the sire be M_1 and M_2. Then the trait values in the offspring can be modelled as

$$y_i = \mu + \beta x_i + e_i \tag{1.1}$$

where y_i is the trait value in offspring i, x_i is an indicator variable whose value is one if an arbitrary marker allele is inherited from the sire and zero otherwise, and e_i is a random error term. A significantly non-zero value for β provides evidence for linkage between the marker and a QTL. The parameter β contains information about the effect size of the QTL, or more accurately the difference between the effects of the two QTL alleles in the sire, but this is confounded with the unknown distance from the QTL to the marker. If the expected difference between the allelic effects at the QTL is a and the recombination fraction between the QTL locus and marker locus is c, then the expected value of β is $(1 - 2c)a$. One problem with this simple analysis is that the allele inherited from the sire is not determinable in the case where the offspring has the same heterozygous genotype as the sire. A typical solution is to ignore all these uninformative individuals (Lynch and Walsh, 1998), but this may result in the loss of a great deal of information. Dentine and Cowan (1990) demonstrate the use of generalized least squares to include uninformative individuals by weighting their contribution using population allele frequencies. The only hurdle to extending this model to multiple half-sib is to realize that the different sires may not be segregating for the same alleles and thus the estimation for marker allele effects needs to be done separately for each sire. This is achieved by fitting the relevant interaction between sire and marker inheritance.

The extension to interval mapping has been demonstrated using both maximum likelihood and regression methods (Knott *et al.*, 1996). As these two methods were found to give similar results, only the regression method will be described due to its direct relationship to the single marker analysis above. Essentially, interval mapping is implemented by calculating the probability that an individual has inherited an arbitrarily chosen haplotype from the sire at each point along the chromosome. This probability is then used for x_i in Equation 1.1. The

calculation of the inheritance probabilities at a particular point in the genome requires only the the inheritance status at flanking informative markers and the recombination distances to these markers (Knott *et al.*, 1996).

An important extension of the half-sib design is the granddaughter design proposed by (Weller *et al.*, 1990). In this design only the male offspring in the half-sib family are considered. However, these also are the sires of further half-sib families providing granddaughters to the original sire, termed the grandsire. The trait values for the sons of the grandsire can then be estimated using the mean of its offsprings' trait values. This provides increased power for the detection of QTL through the increased accuracy of the measured trait value. This design is especially useful in the analysis of milk production characteristics and is widely used in QTL mapping in cattle because the population structure already exists due to the wide use of artificial insemination (AI) and progeny testing.

The extension of this approach of modelling transmission of genotypes from parent to offspring for use in general pedigrees is not practical due to the rapidly increasing number of possible genetic combinations as the pedigree complexity increases (Lynch and Walsh, 1998). A solution to this problem was presented by (Goldgar, 1990), who suggest using a variance component framework in the modelling of QTL. The use of this methods has been further elaborated by a number of authors (Amos, 1994; Almasy and Blangero, 1998; George *et al.*, 2000). The methodology is based around the covariance in trait values for pairs of individuals which may be written as

$$\sigma(y_i, y_j) = G_{ij}\sigma_v^2 + A_{ij}\sigma_u^2 \tag{1.2}$$

where G_{ij} is the estimated identity-by-descent (IBD) proportion for individuals i

and j at the point being examined and A_{ij} is their additive genetic relationship. The variances, σ_v^2 and σ_u^2 represent the amount of the variance in the trait explained by the position of the genome being examined and the remainder of the genome respectively. Under the assumption of multivariate normality of the phenotypic values, a full likelihood model can be formed. In terms of a mixed model framework, the whole pedigree is modelled as

$$\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e} \tag{1.3}$$

where $\mathbf{y}$ is a vector of trait values, μ is the mean of the phenotypes, $\mathbf{Z}$ is a (nxq) matrix relating animals to phenotypes, $\mathbf{u}$ is a vector of additive polygenic effects, $\mathbf{v}$ is vector of additive QTL effects and $\mathbf{e}$ is a residual vector. The vectors $\mathbf{u}$ and $\mathbf{v}$ are assumed to be multivariate normally distributed with zero mean and variances $\mathbf{A}\sigma_u^2$ and $\mathbf{G}\sigma_v^2$ respectively. Evidence for a QTL linked to the chromosomal position being analysed is provide by a significantly non-zero value for σ_v^2. The distribution of likelihood ratio test statistics under the null hypothesis of $\sigma_v^2 = 0$ is a 50:50 mixture of a point mass at zero and a χ_1^2 distribution due to the variance parameter being at the edge of its parameter space under the null hypothesis (Self and Liang, 1987; Stram and Lee, 1994).

Association Mapping: Typical linkage mapping experimental localize the position of the underlying QTL to a region in the order of 20cM. Thus, there are potentially hundreds of genes from which to make the identification of the gene(s) effecting the quantitative trait. In sheep, some progress may be made in reducing the number of genes by using examining the genes in the syntenic regions of the genome in other species that are better annotated, primarily those of cattle and humans. However, this may result in many candidate genes being chosen all of which will require further investigation and, in the worst case

scenario, the actual underlying variant(s) may be missed. Ideally, the identified region needs to be reduced to the magnitude of one or a few centiMorgans.

It has been suggested that the use of linkage disequilibrium (LD) will allow the mapping of QTL which much increased accuracy (e.g. Goddard, 1991; Terwilliger and Weiss, 1998; Haley, 1999). Linkage disequilibrium is the non-random association of alleles at two (or more) loci and is caused by a large number of factors, including genetic drift, population admixture and natural selection. These associations are broken down during meiosis by recombination or, in the case of loci on different chromosomes, the independent assortment of chromosomes. Thus, a marker locus shown to be associated with a quantitative trait is likely to be linked to a QTL. The advantage of mapping using LD instead of linkage is that the while linkage mapping uses information on recombinations within a pedigree to localize the QTL, LD mapping use information from inferred ancestral recombinations that are of a far larger number and thus can localize a QTL with greater accuracy.

The simplest approach to LD mapping involves genotyping the region of interest in a random sample from a population. Then a regression of trait values onto single marker genotypes can be performed. This method can be readily extended to regress on multiple marker genotypes although there is evidence against an increase in precision of the estimated QTL position over single marker methods (Grapes *et al.*, 2004). While this provides a simple study design, there is the potential for false positive results in regions unlinked to any QTL. This is an especially important consideration in livestock populations where large amounts of linkage disequilibrium exist between loci on different chromosomes (Farnir *et al.*, 2000; McRae *et al.*, 2002). Thus the combined use of linkage and linkage disequilibrium is desirable.

A maximum likelihood method for the combined linkage and linkage disequilibrium analysis in half-sib pedigrees has been described by Farnir *et al.* (2002). The method is an extension of that proposed for mapping discrete traits by Terwilliger (1995) that has been adapted for use with large half-sib families. The likelihood is constructed under the assumption of a bi-allelic QTL for which one the the alleles arrived in the population a fixed number generations ago through mutation or admixture. This results in linkage disequilibrium being created between the new QTL allele and the marker alleles in the haplotype it originated on. Thus, in future generations the new QTL allele will be inherited with a greater than random proportion of the marker allele it originated with and it is this excess that is tested for in the maximum likelihood model. This method has been used to map a QTL with a major effect on milk fat content to a 3cM interval (Farnir *et al.*, 2002) and resulted in the subsequent identification and characterization of the underlying polymorphism (Grisart *et al.*, 2002).

Meuwissen and Goddard (2000; 2001) propose a method for linkage disequilibrium mapping in general pedigrees. This involves the use of mixed linear models where a random effect for haplotype similarity, defined later, is fitted. Thus, the framework is very similar to that described above for linkage mapping in general pedigrees and can therefore be extended to include both linkage and linkage disequilibrium information. The similarity of a pair of haplotypes at a particular point of the genome is measured by the number of continuous markers on each side of the point for which the pair have identical alleles. This is the converted to a probability of being IBD at the point of interest using gene-dropping simulations, typically assuming a constant population size of N_e for T generations. The probability that the haplotypes are IBD at the point being tested is measured as the proportion of simulated haplotype pairs showing the same allelic similarities

as the pair of interest that were also IBD at the point being considered. These measures are averaged over a large number of simulation replicates. Thus, an IBD matrix is created using linkage disequilibrium information. This method has been used to fine-map a twinning locus in cattle to a region of less that 1cM (Meuwissen *et al.*, 2002).

1.4 Quantitative Trait Loci Mapping Experiments

A number of QTL mapping studies have been performed in sheep, primarily examining the genetics of traits that are commercially important. This section presents a summary of the current understanding of the genetics of these well studied traits.

Reproduction: The fecundity of a sheep population can be considered as one of its most commercially important traits. This is also one of the most successful areas in sheep genetics in terms of identifying the genes underlying variation between individuals. Several gene with large effect on ovulation rate have been detected and some of these are characterized at the molecular level.

The most well known of these genes was observed segregating in crosses of Booroola Merino (Davis *et al.*, 1982; Piper and Bindon, 1982). This locus, labelled *FecB*, had an autosomal dominant mode of inheritance with each copy increasing ovulation rate by approximately 1.5 times resulting in about one extra lamb per mating (Piper *et al.*, 1985). The *FecB* gene was detected as being linked to two microsatellite markers found on human chromosome 4 (Montgomery *et al.*, 1993). These were subsequently determined to be on sheep chromosome 6 (Montgomery *et al.*, 1994). Almost simultaneously, three groups identified that Booroola sheep carry a mutation in a receptor (*BMPR-1B*) that is expressed in the ovaries (Mulsant *et al.*, 2001; Souza *et al.*, 2001; Wilson *et al.*, 2001).

A second major gene for ovulation rate was identified in the prolific descendants of a Romney ewe (Davis *et al.*, 1991). Segregation analyses demonstrated that this gene, called Inverdale ($FecX^I$), was inherited on the X chromosome as male carriers passed it on to all daughters but no sons. In female carriers of this gene, litter size increases by approximately 0.6. However, ewes homozygous for the mutant allele are infertile (Davis *et al.*, 1995). A second, apparently unrelated, population was discovered to be segregating for an allele, named Hanna ($FecX^H$), which had a similar mode of inheritance. Complementation studies demonstrated these were variants of the same gene (Davis *et al.*, 1995). The gene involved has been identified as *BMP15* (bone morphogenetic protein 15) and the mutations in both the Inverdale and Hanna alleles characterized (Galloway *et al.*, 2000).

A third gene of major effect was identified in Cambirdge and Belcare sheep (Hanrahan, 1991). This was located to sheep chromosome 5 and identified as *GDF9.* (Hanrahan *et al.*, 2004). Both these breeds were also found to be segregating for new allelic variants of the *BMP15* gene (Hanrahan *et al.*, 2004). Other major genes have been mapped to a novel regoin of the X chromosome (Woodlands (*FecX2*); Davis *et al.*, 2001) and chromosome 11 (Lacune (*FecL*); Lecerf *et al.*, 2002). Other populations have also been identified as segregating for major genes (Davis, 2005). The large number of genes the have major effects on ovulation rate is interesting in itself. However, the lack of genes or chromosomal regions identified as having small effects, as is typical with most other quantitative traits, can be attributed to the focus on identifying the genes with obvious large effects. Thus, general conclusions about the architecture of the genetics of ovulation rate should be avoided until more complete information is available.

Carcass Traits: Carcass traits, including body weight, meat content and quality, have received less attention than reproductive traits as they provide a less direct impact on profit (Cockett *et al.*, 2005). Large differences in meat traits are observable between sheep breeds suggesting substantial improvements could be made through genetic selection (Thompson and Ball, 1997).

Several major loci for muscle traits have been identified in sheep. The first of these, the *callipyge* locus, causes pronounced hypertrophy of the fast twitch muscle fibres, primarily those on the pelvic limb. This results in carcasses that are with increased and leaner meat (Koohmaraie *et al.*, 1995; Jackson *et al.*, 1997). The *callipyge* locus shows a novel mode of inheritance called polar overdominance, where it is only expressed in heterozygous individuals inheriting their mutation from their sire (Cockett *et al.*, 1996; Freking *et al.*, 1998). This gene was mapped to the distal region of chromosome 18 (Cockett *et al.*, 1994) and subsequently fine-mapped to a region of 4.6cM (Segers *et al.*, 2000) and then a 450kb region (Berghmans *et al.*, 2001). A causal SNP has been identified within an imprinted gene cluster (Freking *et al.*, 2002; Smit *et al.*, 2003), although its exact function is still unknown.

A second major locus for muscle mass has been located to the distal end of chromosome 18 (Nicoll *et al.*, 1998). This locus, called *Carwell* or *rib-eye muscling*, causes hypermuscling in the longissimus, or rib-eye, muscle in sheep. While the location of the locus and its effect on muscle mass show similarities with the *callipyge* locus, the mode of inheritance does not include any parent-of-origin effects (Jopson *et al.*, 2001). Thus, the relationship between these loci is unclear at present.

A further major locus for muscular hypertrophy has been identified in Texel sheep, which is referred to as the double muscling locus due to its phenotypic similarity to double muscling in cattle. As the *myostatin* gene has been identified as the cause of double muscling in cattle (Arnold *et al.*, 2001), this provides a potential candidate locus for this trait in sheep. Indeed, the gene underlying the double muscling in sheep has been map to the region on chromosome 2 that includes the *myostatin* gene (Marcq *et al.*, 1999). However, the sequencing of the coding region of the *myostatin* gene has not identified the causal mutation (Marcq *et al.*, 1998).

Other loci showing effects on growth traits include the *transferrin* gene on chromosome 1 (Kmiec, 1999a; Kmiec, 1999b) and the MHC locus on chromosome 20 (Paterson *et al.*, 1998; Bot *et al.*, 2004). Further regions effecting muscle and fat content have been detected on chromosome 3, 4 and 6 (Walling *et al.*, 2004).

Disease Resistance: The genetic of resistance of sheep to a variety of diseases has been examined. Perhaps the most studied is the resistance to internal parasites due to their negative effect of production traits. It is well known that there exists a large variance in parasite resistance between breed (Beh and Maddox, 1996). Within breeds, heritability estimates average around 0.3 (McEwan *et al.*, 1994; Douch *et al.*, 1995a; Douch *et al.*, 1995b; Woolaston and Piper, 1996). A whole genome scan by Beh *et al.* (2002) did not find any QTL significant at the genome-wide level despite the reasonable sample size of 1029 sheep in six half-sib families. However, a QTL significant at the chromosome-wide level was found on chromosome 6 and point-significance was obtained on chromosomes 1 (distal), 3, 11 and 12. The confidence interval for the QTL on chromosome 3 covered the region near the sheep interferon-γ gene

is located. A locus for parasite resistance has been mapped to within 7cM in this region (Paterson *et al.*, 2001). Associations with parasite resistance have also been observed on chromosome 20 around the major histocompatibility locus (Schwaiger *et al.*, 1995; Janssen *et al.*, 2002).

Resistance to scrapie has received considerable attention recently, with many countries beginning to implement breeding schemes to reduce overall susceptibility. The primary genetic influence on scrapie susceptibility is the genotype at the prion protein gene, PnP, on chromosome 13 (reviewed in Tranulis, 2002; Baylis and Goldmann, 2004). Another disease for which the genetics are beginning to be studied is facial eczema. Candidate gene studies have been performed specifically looking at antioxidants, with the gene catalase on chromosome 15 showing some association (Phua *et al.*, 1999). A resource population with highly divergent lines has been developed for QTL mapping (Morris *et al.*, 2004).

Other Traits: Several studies have been performed, or are in the process of being performed, on the genetics of other commercially important traits, especially in the areas of wool and milk production. These are only briefly considered here as no regions effecting non-discrete traits have been replicated or had their underlying polymorphism discovered. The genetics of Mendelian traits in wool production have been well studied, especially the in the areas of the colour of skin and wool fibre (Sponenberg, 1997). Several studies have identified regions containing putative QTL for wool production and quality (reviewed in Purvis and Franklin, 2005). The genetics of milk production is only beginning to studied in sheep, with preliminary results from a QTL scan and a review of other resource populations being given in Barillet *et al.* (2005).

1.5 Thesis Overview

This thesis aims to increase the understanding of quantitative trait variation in both domestic and feral sheep populations. This is approached both directly through linkage analysis and indirectly through the examination of methodology for use in linkage analysis and performing the groundwork for the design of linkage disequilibrium studies in a feral sheep population.

The following chapters of this thesis can roughly be split into two sections. The first involves the a directed linkage scan for loci involved in body weight and carcass composition in a complex commercial Charollais sheep population using the variance components approach (Chapter 2). The analysis of this data set demonstrated an inconsistency with the published linkage map, which is further characterized in Chapter 3. The QTL mapping also raises questions about the methodology for evaluating significance thresholds with general pedigrees. Chapter 4 examines the properties of one possible solution to this problem.

The second section examines the population dynamics of a well studied feral sheep population. In Chapter 5 a statistical framework is developed for the modelling of all major aspects of the life cycle of the sheep. This is then used as the basis for a simulation study that is used to predict the amount of linkage disequilibrium in the population (Chapter 6) and the effective population size of the population (Chapter 7). The thesis is concluded in Chapter 8 with a summary and general discussion.

2 Mapping of Multiple Quantitative Trait Loci for Growth and Carcass Traits in a Complex Commercial Sheep Pedigree

2.1 Introduction

The last decade has seen a large number of experiments to discover quantitative trait loci (QTL) of commercial benefit segregating in livestock populations. Generally, these experiments have met with success in terms of identifying regions containing QTL, but the variants underlying the discovered QTL have been discovered in very few cases (e.g. Grobet *et al.*, 1998; Kim *et al.*, 2000; Wilson *et al.*, 2001; Grisart *et al.*, 2002). Therefore, direct evaluation of the importance of the discovered QTL in (other) commercial populations is usually not possible. Instead, further QTL experiments need to be performed on candidate chromosomal regions in the commercial populations of interest (e.g. Nagamine *et al.*, 2003; de Koning *et al.*, 2003; Walling *et al.*, 2004).

Animals in national improvement schemes provide opportunities for the evaluation of the importance of QTL as these animals typically have good pedigree and phenotypic information. The widespread use of artificial insemination in these populations typically creates large half-sib families, thus potentially providing study designs similar to those used in experimental populations. However, such designs are not necessarily optimal for the confirmation of a QTL segregating in a population of interest. For the genotyping of a fixed number of individuals, there is a trade off between the probability of having a QTL segregating in a half-sib family and the power of detection of the QTL (Weller *et al.*, 1990). This potential pitfall can be reduced by the use of a large number of related families, where the relationships between families can be used to mitigate some of the power lost due to increasing

the probability of observing QTL segregation (Williams *et al.*, 1997; Slate *et al.*, 1999).

The analysis of complex pedigrees suffers from the computationally demanding nature of the calculations involved. However, their use is becoming increasingly widespread in populations where experimental intervention is not practical. George *et al.* (2000) demonstrated a two-step procedure for variance component interval-mapping in complex pedigrees that contain missing marker information. Firstly, the proportion of genes identical-by-descent (IBD) between all individuals is estimated at each chromosomal location using a Markov chain Monte Carlo (MCMC) sampling procedure. Then the contribution of the chromosomal location to the phenotypic variance is assessed using restricted maximum likelihood (REML). This approach has been used to map QTL in humans (Visscher *et al.*, 1999), wild deer populations (Slate *et al.*, 2002b) and commercial pig populations (de Koning *et al.*, 2003). Here, this methodology is used to investigate growth and carcass traits in a complex pedigree from a commercial sheep population.

2.2 Materials and Methods

Animals: A selection of 570 related sheep from the United Kingdom Charollais Sire Referencing Scheme formed the basis of the pedigree to be analysed. Animals were chosen on the basis of being the descendants of five widely used sires. Emphasis was placed on identifying sheep that derived from mating between the descendants of these sires. A total 406 sheep from this pedigree were genotyped, with the majority of ungenotyped sheep being ancestors in the pedigree. This pedigree contained evidence of inbreeding with 70 loops in total. Each animal was weighed at 8 weeks of age (EWW) and at ultrasonic scanning (SCW) undertaken at approximately 20 weeks of age. At scanning, muscle

depth (MUS) and fat depth (FAT) at the third lumbar vertebra were recorded. Both muscle and fat traits were also analysed following correction for live weight giving two further traits (MWT and FWT, respectively).

Selection of Chromosomes: Chromosomes to be genotyped were selected based on previous studies in sheep or other livestock species showing the likely presence of major QTL for traits related to the growth and carcass traits being analysed. Chromosome 1 has been shown to contain growth effects around the transferrin gene (Kmiec, 1999a; Kmiec, 1999b). Chromosome 2 contains muscling effects near the myostatin gene (Marcq *et al.*, 1998; Broad *et al.*, 2000; Walling *et al.*, 2001). Chromosome 3 is syntenic to the region around insulin-like growth factor-1 (IGF-1) that shows growth effects in cattle (Moody *et al.*, 1996; Stone *et al.*, 1999; Casas *et al.*, 2000; Machado *et al.*, 2003). Chromosome 18 contains the callipyge gene (Cockett *et al.*, 1994; Freking *et al.*, 2002) and the Carwell rib eye muscling locus (Nicoll *et al.*, 1998), which are possibly allelic. Chromosome 20 contains the MHC locus that has been shown to affect growth traits in sheep (Paterson *et al.*, 1998; Bot *et al.*, 2004) as well as cattle (Elo *et al.*, 1999) and pigs (Jung *et al.*, 1989).

Genotyping: The five chromosomes were genotyped at a total of 69 markers chosen for having a polymorphic information content (PIC) greater than 0.6. This gave an average marker spacing of approximately 17cM. Entire chromosome lengths were covered, allowing confidence intervals for QTL position to be constructed and compared to the candidate regions. Marker order was checked against that given by Maddox *et al.* (2001) using Cri-Map (Green *et al.*, 1990). Updated map distances from the map given in Maddox *et al.* (2001) were used in the analyses due to the comprehensive nature of their data set (Maddox, 2003).

Preparation of Phenotypic Data: The phenotypic measurements were corrected for known fixed effects and covariates using a dataset of approximately 42000 sheep which had been measured in the Charollais Sire Referencing Scheme. Pre-correction of phenotypes was chosen over the simultaneous analysis of all effects as the latter option would require the inversion of a 42000 by 42000 matrix of IBD probabilities at each chromosomal position examined in the variance component analysis. More accurate estimates of fixed effects are obtained by pre-correcting the phenotypes using the larger data set and the smaller standard errors obtained have no effect on further results as only residuals from the models are used in the QTL mapping. All traits were corrected for the sex of the sheep, year of birth, the flock that they were raised in, their birth-rearing rank and the linear effect of the age of their mother. SCW, MUS and FAT were further corrected for the linear effect of age (in days) at the time of scanning, while MWT and FWT were corrected instead for the linear effect of weight at scanning.. All models were fitted using ASREML (Gilmour *et al.*, 2002) and residuals were checked for normality.

Heritability Estimation: Heritabilities of the adjusted traits were estimated using the animal model approach (Lynch and Walsh, 1998). Briefly, this model is written in matrix notation as

$$\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e} \tag{2.1}$$

where $\mathbf{y}$ is an (nx1) vector of phenotypes, μ is the mean of the phenotypes, $\mathbf{Z}$ is an (nxq) incidence matrix relating animals to phenotypes, $\mathbf{u}$ is an (qx1) vector of additive polygenic effects and $\mathbf{e}$ is the (nx1 residual vector. The random effects $\mathbf{u}$ and $\mathbf{e}$ are assumed to be uncorrelated and distributed as multivariate normal densities as follows: $\mathbf{u} \sim N_q(0, \mathbf{A}\sigma_u^2)$ and $\mathbf{e} \sim N_n(0, \mathbf{I}\sigma_e^2)$ where $\mathbf{A}$ is the standard

additive genetic relationship matrix and **I** is the identity matrix. The estimation of the variance components was done using ASREML (Gilmour *et al.*, 2002) and the heritability of a trait, h^2 was calculated as

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \tag{2.2}$$

Pedigree Error Assessment: The probability of an incorrectly assigned parent-offspring trio having inconsistent genotypes at a marker locus was calculated as

$$Q_l = 1 - 2S_2 + S_3 + 2S_4 - 2S_2 - 3S_5 + 3S_3S_2 \tag{2.3}$$

where

$$S_t = \sum_i p_i^t \tag{2.4}$$

and p_i is the frequency of allele i at the locus (Dodds *et al.*, 1996). The distribution of the number of inconsistent genotypes expected for an incorrectly assigned parent-offspring trio was approximated using a binomial distribution with $n = 69$ and probability $Q = 0.4938$ being the average of Q_l over all loci l. From this approximation, 99.5% of all pedigree errors involving parent-offspring trios will have greater than 23 mismatches. A similar calculation was performed for parent-offspring pairs using

$$Q_l = 1 - 4S_2 + 4S_3 - 3S_4 + 2S_2^2 \tag{2.5}$$

(Dodds *et al.*, 1996) which gave an average $Q = 0.3289$. Using the binomial approximation, 99.5% of all parent-offspring pairs will have greater than 12 inconsistencies. Offspring in pairs with more than 12 inconsistencies had their parent set to unknown in the pedigree. When parent-offspring trios had more than 23 inconsistencies, each parent was tested individually to check if the inconsistency could be assigned to one parent and the appropriate parent(s) was

set to unknown in the offsprings pedigree entry. All other inconsistent parent-offspring genotypes were removed. More complex genotype inconsistencies were removed using PedCheck (O'Connell and Weeks, 1998). The above thresholds cannot be applied to these complex inconsistencies but cases that showed repetitive inconsistencies of the same type (e.g. more than four alleles in a full-sib family with two ungenotyped parents) occurred in only a few clear-cut cases.

Half-sib Analysis: The genotyped pedigree contained a half-sib family with 51 progeny. This was analysed with a regression based interval mapping method, developed from the method of Knott *et al.* (1996), using QTL Express (Seaton *et al.*, 2002). Briefly, the corrected phenotype is regressed upon the conditional probability that a particular haplotype is inherited from the sire. The test statistic is an F ratio with 1 and $n-2$ degrees of freedom where n is the size of the half-sib family. To allow comparison with further analyses, the F statistic was transformed into a likelihood-ratio statistic by

$$LRT = \begin{cases} (n-1)\ln\left(\frac{n-2}{n-1} + \frac{F}{n-1}\right) - \ln(F); & F > 1 \\ 0; & F \leq 1 \end{cases} \tag{2.6}$$

(Baret *et al.*, 1998). Chromosome-wide significance of possible QTL were determined by permutation testing (Churchill and Doerge, 1994) and confidence intervals for QTL location were constructed by bootstrap analysis (Visscher *et al.*, 1996). Other half-sib families in the pedigree were considered too small for the detection of QTL at a moderate power.

Variance Component Analysis: The complete 570 sheep pedigree was analysed using the two-step approach proposed by George *et al.* (2000). Firstly identity-by-descent (IBD) coefficients between all individuals were determined using LOKI (Heath, 1997) at 1-cM intervals along the chromosomes. IBD coefficients

were calculated from the mean of samples taken every 10 iterations after a 1000-iteration dememorization period. Chromosomes 1, 2, 18 and 20 required 10,000 samples to obtain good concordance between repeated runs of LOKI and chromosome 3 required 100,000. The QTL effect at each chromosomal location was modelled as

$$\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e} \tag{2.7}$$

where $\mathbf{y}$, $\mathbf{Z}$, $\mathbf{u}$ and $\mathbf{e}$ are as defined in the estimation of heritability and $\mathbf{v}$ is a (qx1) vector of additive QTL effects. The distribution of $\mathbf{v}$ is assumed to be $\mathbf{v} \sim N_q(0, \mathbf{G}\sigma_v^2)$ where $\mathbf{G}$ is the (qxq) (co)variance matrix for the additive QTL effects, represented by the proportion of alleles IBD. At each step along the chromosome the variance explained by the QTL effect is tested for significance by

$$LRT = -2\ln(L_0 - L_1) \tag{2.8}$$

where L_1 is the log-likelihood of the model including the QTL effect and L_0 is the log-likelihood without the QTL effect. Likelihoods were calculated using ASREML (Gilmour *et al.*, 2002). For a single chromosome location, the likelihood ratio statistic is distributed as a 50:50 mixture of a point mass at 0 and a χ_1^2 distribution. Due to the computationally demanding nature of variance component QTL mapping, the chromosome-wide significance of the QTL effect cannot be directly computed. However, simulation studies show the chromosome-wide test statistic to be distributed between a χ_1^2 distribution and a χ_2^2 distribution (Xu and Atchley, 1995; Grignola *et al.*, 1996). Here, a χ_2^2 distribution was assumed for stringency.

2.3 Results

The trait data are summarized in Tables 2.1 and 2.2. From the differences between the standard deviation for the raw and corrected data, known

Table 2.1: Summary of measured trait data.

Trait[a]	Mean	Standard Deviation	Residual Standard Deviation[b]	Heritability (s.e.)
EWW	22.34	4.68	3.70	0.25 (0.012)
SCW	50.73	10.14	5.78	0.27 (0.012)
FAT	3.61	1.77	1.29	0.25 (0.013)
FWT			1.12	0.27 (0.013)
MUS	28.17	3.57	2.60	0.25 (0.013)
MWT			2.21	0.31 (0.014)

[a] See **Preparation of Phenotypic Data** for trait definitions.

[b] Standard deviation of the phenotypic residual values after correcting for fixed effects and covariates included in the model

environmental effects are estimated to account for between 37% and 68% of the variation in trait values demonstrating the importance of the pre-correction of trait data. All traits showed a significant heritability with lowest estimated heritability being 0.25 for FAT and the largest 0.31 for MWT. As expected from the nature of the traits being analysed, there were significant phenotypic correlations among corrected traits. The largest of these correlations occurred between the pairs of traits corrected from the same raw phenotype (i.e. FAT with FWT and MUS with MWT) and the two live weight traits (EWW and SCW). The lack of significant correlation between weight-corrected traits (FWT, MWT) and SCW is expected, as SCW was included as a covariate in the correction of these traits.

Analysis of genotyped animals revealed 16 sheep with inconsistent parental information. If this sub-population is representative of the population as a whole, this indicates that up to 6.5% of parent-offspring pairs in the United Kingdom Charollais Sire Referencing Scheme are incorrectly assigned. The map order of

Table 2.2: Phenotypic correlation matrix of corrected trait data.

	Trait				
Trait[a]	**EWW**	**SCW**	**FAT**	**FWT**	**MUS**
SCW	0.650				
FAT	0.279	0.495			
FWT	-0.046	-0.001^{ns}	0.868		
MUS	0.311	0.532	0.327	0.072	
MWT	-0.034	0.009^{ns}	0.079	0.086	0.851

[a] See **Preparation of Phenotypic Data** for trait definitions.
$^{ns}p > 0.05$; $p < 0.001$ for all other pairs

the markers used in the Charollais sheep population agreed with the published international mapping flock (Maddox *et al.*, 2001) except for a small region of chromosome 1 where the ordering of markers MCM137 (232.6cM) and BM6506 (234.8cM) relative to their flanking markers was significantly different. Due to these markers being tightly linked, the order used in the analysis had little effect on the QTL profiles (data not shown) and as such the map order given in Maddox *et al.* (2001) was used for consistency. This marker order inconsistency is examined in more detail in the following chapter.

Half-sib Analysis: Chromosome 1 contained QTL significant at the 5% chromosome-wide level for the two weight traits, EWW and SCW. The information content for the half-sib family across chromosome 1 is displayed in Figure 2.1. The lack of information at about 150cM is a result of using markers chosen for overall polymorphic information content (PIC) rather than being heterozygous in the half-sib family. Approximately 70% (16 out of 23) of markers genotyped on chromosome 1 were heterozygous in the half-sib family providing good agreement with the selection criteria that required a PIC greater than 0.6.

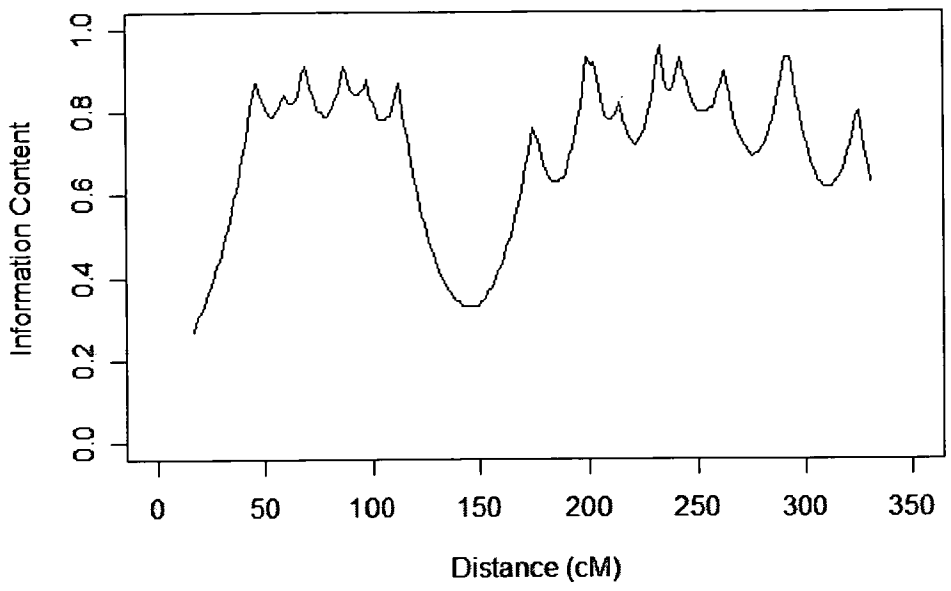

Figure 2.1: Information content for the half-sib family QTL analysis on chromosome 1. The information content shows a large drop at around 150 cM due to markers being chosen for their average information content rather than heterozygosity in the common sire.

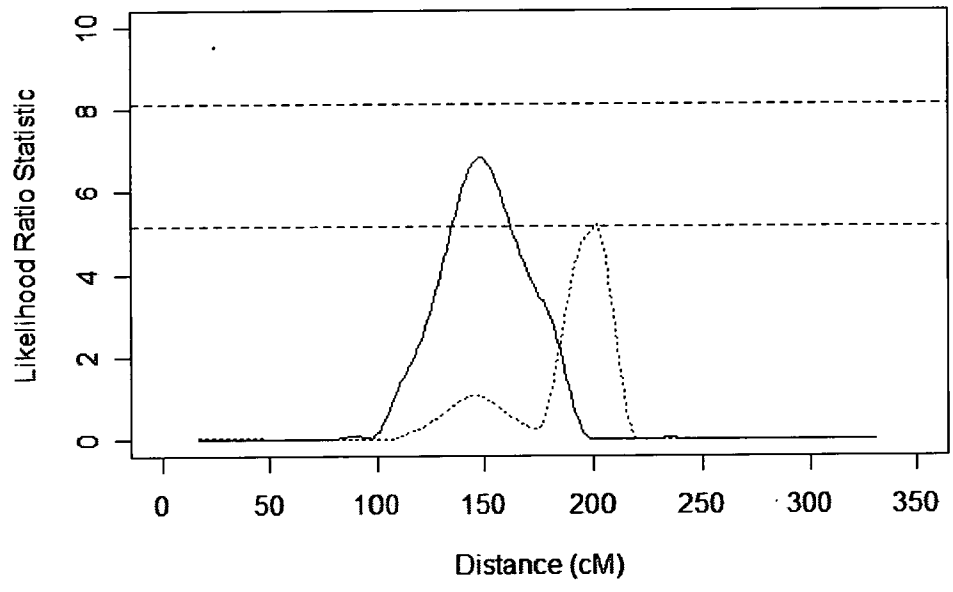

Figure 2.2: QTL profiles for the significant traits from the analysis of the half-sib family on chromosome 1. Both live weight at ultrasonic scanning at approximately 20 weeks of age (SCW, solid line) and 8-week weight (EWW, dotted line) were significant at the 5% chromosome-wide level. Dashed lines indicate 1% and 5% permutation significance thresholds.

The QTL profiles for SCW and EWW on chromosome 1 are shown in Figure 2.2. The peak for SCW occurs at 148cM (between heterozygous markers MCM58 at 112.9cM and ILSTS04 at 175.1cM) in the middle of the area with low information content. The 95% bootstrap confidence interval for QTL position is 70 to 239cM. The effect of allelic substitution is 1.21 residual standard deviations (s.e. = 0.36). The QTL for EWW is estimated to be at 202cM (between CSSM04 at 199.8cM and BMS4000 at 203.0cM) with a 95% confidence interval of 24 to 291cM. This allele substitution effect of this QTL is 0.56 residual standard deviations (s.e. = 0.19). A further significant QTL effect was observed on chromosome 18 for SCW (F = 5.28, $p < 0.05$ chromosome-wide). However, only two markers on this chromosome were heterozygous in the common sire (MCMA25 at 96.5cM and OY5 at 118.0cM), so the localization of QTL peak is unreliable and gives a 95% bootstrap confidence interval spanning the entire chromosome.

Variance Component Analysis: A total of seven trait by chromosome combinations reached the approximate 5% chromosome-wide significance level in the variance component analyses, two on each of chromosomes 1 and 2 and three on chromosome 3 (see Figures 2.3 to 2.5). The majority of these QTL were for the two traits derived from fat depth measurements with QTL for FAT on chromosomes 1, 2 and 3 and QTL for FWT on chromosomes 2 and 3. The QTL for FAT and FWT on chromosomes 2 and 3 map to very similar positions, 89cM (1 LOD support interval = 74 to 115cM) and 86cM (71 to 109cM) for FAT and FWT respectively on chromosome 2 and 36cM (0 to 63cM) for FAT and 34cM (0 to 54cM) for FWT on chromosome 3, indicating these are probably the same QTL. The QTL for FAT on chromosome 1 maps to 81cM (62 to 96cM). The remaining QTL were for SCW on chromosome 1 at 276cM (258 to 315cM) and at 32cM (0 to 73cM) on chromosome 3.

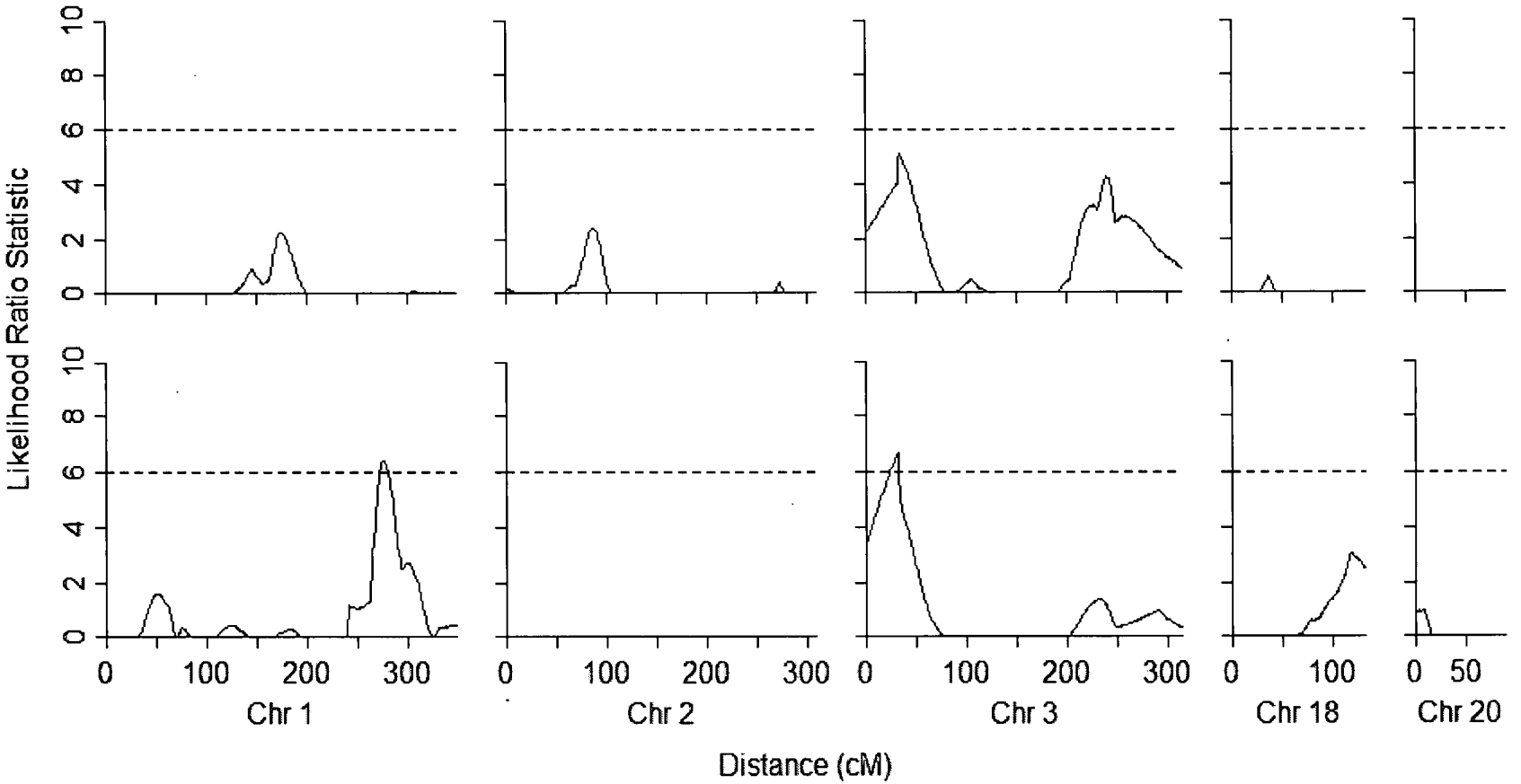

Figure 2.3: QTL profiles from the variance component analysis of live weight traits, 8-week weight (EWW, top) and weight at ultrasonic scanning at approximately 20 weeks of age (SCW, bottom).

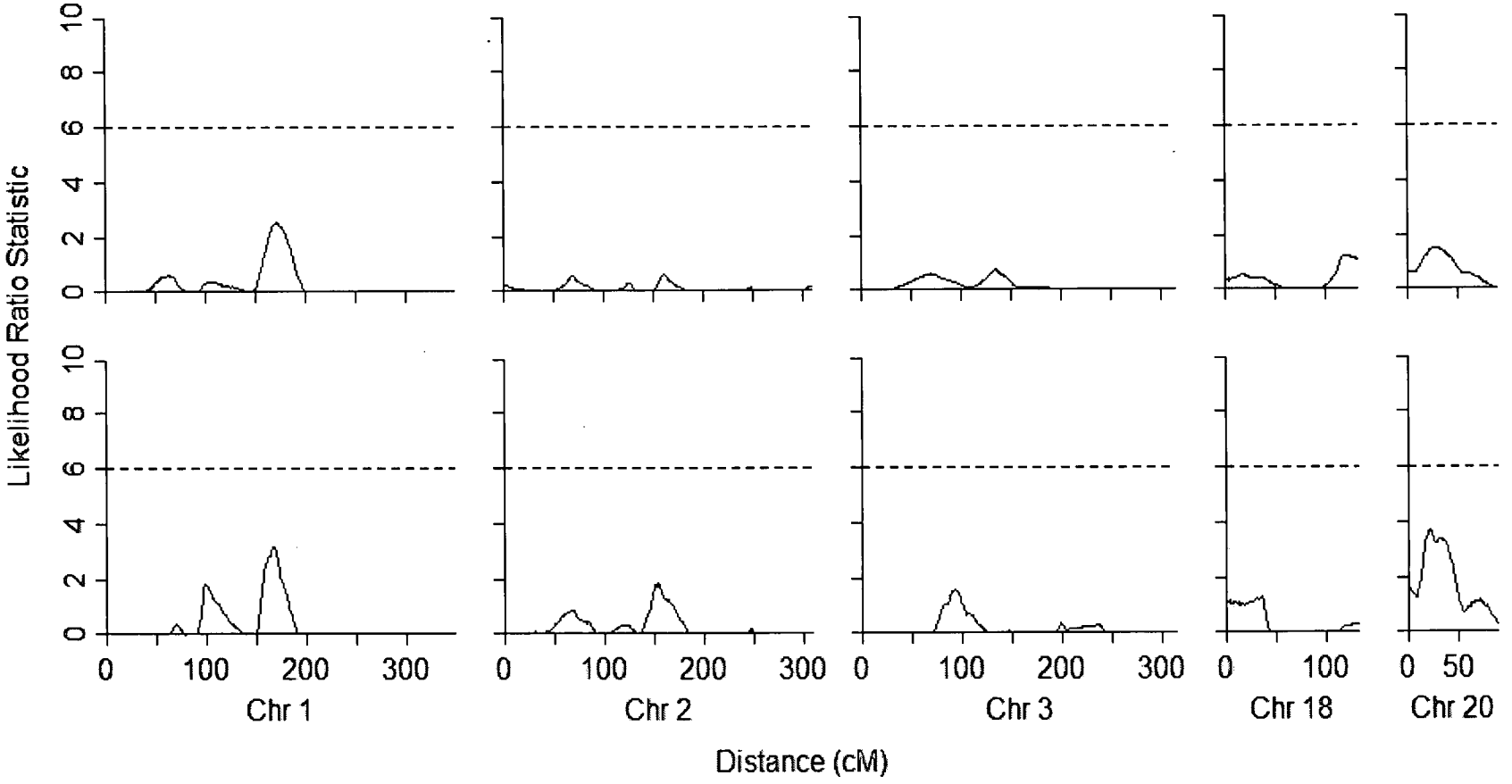

Figure 2.4: QTL profiles from the variance component analysis of muscle traits, muscle depth at the third lumbar vertebrae corrected for age at scanning (MUS, top) and weight at scanning (MWT, bottom).

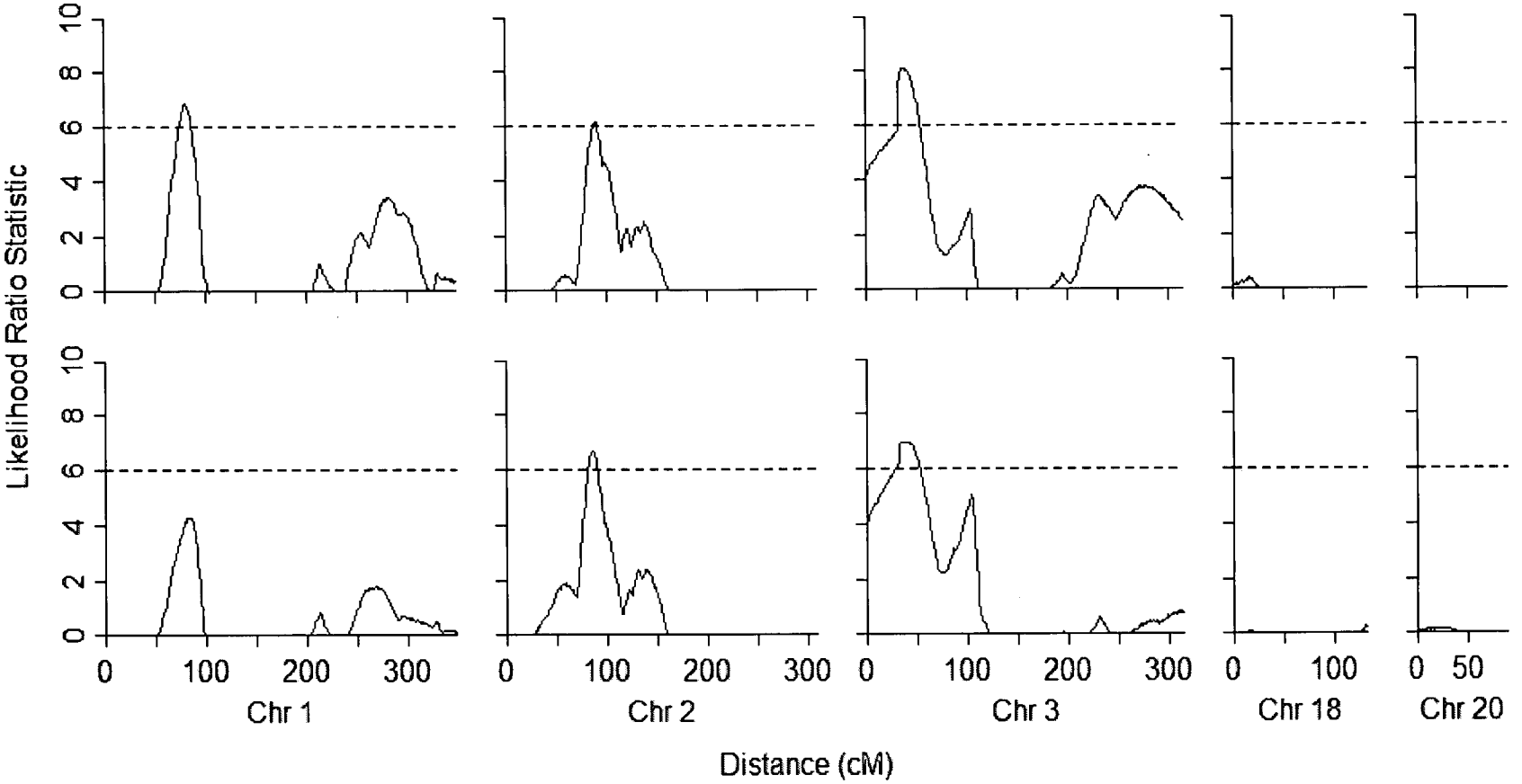

Figure 2.5: QTL profiles from the variance component analysis of fat traits, fat depth at the third lumbar vertebrae corrected for age at scanning (FAT, top) and weight at scanning (FWT, bottom).

The estimates of the proportion of additive genetic variance explained by each QTL ranged from approximately 0.60 to 0.90, giving the total proportion of the additive genetic variance explained by the QTL detected for each of these traits of greater than 1.00. This upwardly biased nature of the estimates of effects determined at the QTL peak has previously been observed by others using this QTL mapping method (George *et al.*, 2000; Göring *et al.*, 2001). The biased nature of QTL effect estimates is a general problem due to the estimation of effects at the QTL peak where the estimate is the greatest and is particularly strong when the power to detect the QTL is low (Lynch and Walsh, 1998). However, as the nature of the bias is consistent among QTL, the relative proportion of variance explained by each QTL can be estimated by including multiple QTL effects in the variance component model. In this way the proportion of the total additive genetic variance explained by the QTL for FAT on chromosomes 1, 2, and 3 were estimated as 0.429, 0.207 and 0.364 respectively. The QTL for FWT on chromosomes 2 and 3 explained 0.560 and 0.440 of the genetic variance respectively and the QTL for SCW explained 0.571 and 0.429 on chromosomes 1 and 3 respectively. As these proportions sum to 1.00, it follows that the remaining polygenic variance was estimated as zero in all cases. This is not surprising given how large QTL effects were when estimated individually. Multiple QTL models also allow for the testing of the significance of each QTL in the presence of the others. Only the QTL for FAT on chromosome 2 fell below significance at a nominal 5% level indicating this QTL did not explain a significant amount of trait variation given the other QTL. However, this is likely to be due to a lack of power for the detection of three QTL simultaneously as the QTL for FWT in the same region remains significant in the presence of the other FWT QTL. As such the QTL for FAT on chromosome 2 should not be discounted as a false positive.

2.4 Discussion

The investigation of the presence of previously discovered QTL in a variety of commercial populations is an important step in the understanding of the commercial relevance of QTL discovered in experimental studies. This approach is facilitated by the use of genetic improvement schemes in which large numbers of animals are routinely produced and have relevant data collected. The pedigrees available from such schemes provide an inexpensive resource for QTL mapping. Also, any significant results are immediately applicable for marker assisted selection schemes.

Typically, with a large amount of pedigree and phenotype data available, the major constraint in the detection of QTL in commercial populations will be the cost of genotyping. This can be reduced by only considering chromosomes with experimental evidence of QTL for similar traits in other populations. However, the appropriate study pedigree needs to be selected from the available data in order to maximize the power of QTL detection. The use of half-sib families is an attractive option given the widely available resources for the analysis of such data due to its frequent use in experimental populations. However, the detection of QTL in such designs requires the QTL to be segregating in the common parent. In the best-case scenario where the QTL has allele frequencies of 0.5 for each state, approximately 50% of animals will be segregating for the QTL (under idealized population conditions). Thus five families will be needed to provide a greater than 95% chance of having at least one parent segregating for the QTL. However, if the QTL alleles have frequencies of 0.1 and 0.9, a case that is more realistic for highly selected traits, 17 families are required for a similar probability of QTL segregation. With a fixed genotyping budget, increasing the number of families rapidly decreases the power to detect the segregation of QTL in each family. This problem can be potentially reduced by the use of related

families where the additional relationships can be used to provide an increase in power to detect QTL (Williams *et al.*, 1997; Slate *et al.*, 1999). Although the analysis of complex pedigrees is comparatively very computationally demanding, the use of such designs is becoming possible even with moderately large pedigrees containing substantial inbreeding.

Chromosome 1 was chosen because of the presence of the transferrin gene at 299cM, which has been shown to be associated with growth effects (Kmiec, 1999a; Kmiec, 1999b). Four QTL were found to be segregating on chromosome 1 of the UK Charollais sheep population, two with the half-sib analysis and two by variance component analysis. The halfsib analysis found significant QTL for SCW and EWW at 148 and 202cM respectively. The power of the half-sib analysis to distinguish between one and two QTL was reduced due to the selection of markers for information content in the whole pedigree. This resulted in a region of low information content in the centre of chromosome 1 for the half-sib analysis and large overlapping confidence intervals for SCW and EWW whose peaks occurred around this region. However, both confidence intervals for the QTL position exclude the transferrin gene. The variance component analysis found QTL for FAT and SCW with peaks at 81 and 276cM respectively. The one LOD support interval for the SCW QTL overlaps the transferrin gene indicating this as a potential candidate gene. The one LOD support interval for the SCW QTL detected in the variance component analysis does not overlap the 95% confidence interval for the SCW QTL detected in the halfsib analysis. However, one LOD support intervals have been shown to be smaller than their respective 95% confidence intervals (Van Ooijen, 1992; Mangin *et al.*, 1994) and the more appropriate two LOD support interval for SCW covers the entire chromosome. Thus there is not enough evidence from this study to conclude that there is more than one QTL for SCW on chromosome

1. Discrepancies between result from the half-sib analysis and the variance component analysis such as observed here appear to be common even when analysing half-sib families using both methods (Slate *et al.*, 2002b; de Koning *et al.*, 2003). Here, QTL found using the variance component analysis but not in the half-sib analysis are likely to be due to the QTL alleles not segregating in the common parent of the half-sib family.

Significant QTL were found on chromosome 2 for the two traits derived from fat depth measurements, FAT and FWT. Given the high correlation between these traits and the similar QTL profiles, it is likely that there is one underlying QTL affecting both traits. Chromosome 2 was chosen for the mounting evidence of one or several QTL for carcass composition segregating around the myostatin locus (Marcq *et al.*, 1998; Broad *et al.*, 2000; Walling *et al.*, 2001). However, the regions covered by the one LOD support intervals for both traits are approximately 90cM proximal from the region around myostatin in which growth effects have been observed. Three QTL were found with the variance component analysis of chromosome 3, one for each of FAT, FWT and SCW. The confidence intervals for these traits cover the proximal 70cM of chromosome 3, again excluding the candidate IGF-1 locus at 227cM. No significant QTL were found on chromosomes 18 and 20, apart from a significant effect for SCW on chromosome 18 in the half-sib analysis which provided little information due to lack of segregating markers in the common parent.

The use of a candidate region approach to detect QTL in the Charollais sheep population has not been successful in the detection of QTL previously observed in other livestock populations. Of the nine significant QTL detected in this study, only one maps to a candidate region (SCW on chromosome 1). This is currently being confirmed in a larger sample from the Charollais sheep population before

marker assisted selection is implemented. Confirmation studies are essential before any of the QTL detected in this study can be used in a marker assisted selection scheme. The upwardly biased estimates for QTL size suggest selecting only on the marker information with no additional phenotypic selection as the residual additive variance is estimated to be zero. However, unbiased estimates of QTL sizes can be obtained by estimating at the current QTL estimate position in the verification study population (Lande and Thompson, 1990). The discovery of QTL in regions not generally recognized as important for the traits analysed indicates that the understanding of quantitative genetics of important traits in a variety of commercial populations needs to be improved. It is realistic to assume that QTL of major effect on the trait undergoing selection are likely to have become close to fixation through the intense selection based on phenotype (or estimated breeding value) that occurs in commercial populations. Thus, QTL with an estimated smaller effect in experimental populations using wide crosses may be more commercially important than those of large effect.

3 Examination of a Region Showing Linkage Map Discrepancies Across Sheep Breeds

3.1 Introduction

The advent of molecular markers has seen the construction of linkage maps for a variety of commercially important livestock species including cattle (Barendse *et al.*, 1997; Kappes *et al.*, 1997; Ihara *et al.*, 2004), chicken (Groenen *et al.*, 2000; Schmid *et al.*, 2000), deer (Slate *et al.*, 2002a), pig (Archibald *et al.*, 1995; Rohrer *et al.*, 1996) and sheep (de Gortari *et al.*, 1998; Maddox *et al.*, 2001). The primary purpose of such linkage maps is to provide a framework to locate genes of commercial importance, either through linkage mapping or comparative analysis. In a typical linkage study, the published linkage map is used, rather than constructing separate linkage maps for each population studied, as this allows for comparisons across studies and removes the need to publish similar linkage maps for each population studied. This approach of using published linkage maps is also justifiable given the more comprehensive coverage of the genome in the mapping populations.

The use of linkage maps in breeds of animals other than those that were used in the construction of the published map requires recognition of the limitations of the methods used in their construction. The map distances provided depend on the mapping function used and are an average over the pedigree(s) investigated, thus averaging any variation between individuals. As some of the variation in recombination has been shown to be heritable (Kong *et al.*, 2004), it is likely that differences in map distances exist between populations. Such differences have been observed between ethnic groups in human populations (Jorgenson *et al.*, 2005). However, differences between the actual map distances and those used in linkage analysis have little effect in practice (Dodds *et al.*, 2004) and are

difficult to detect statistically.

The map order is more fundamentally important in gene mapping, especially when trying to physically locate the underlying polymorphism controlling the trait of interest. As such, the map order is generally the primary focus of linkage map construction with map distances being estimated only as a secondary consequence. This is evident in the published maps that, in general, provide a framework map where the likelihood of the order given is greater than the likelihood of any other order by at least some predefined threshold, but provide no standard errors on distances. A typical threshold of three LOD (logarithm of the odds) is based on a recommendation given by Morton (1955) for a correction for the multiple testing involved in linkage analysis. Given the latest linkage maps for the major livestock species now contain more than 1000 loci, a situation clearly not considered by Morton, the continued use of this threshold is likely to result in some errors in the estimated marker order.

Despite these limitations in linkage map construction, differences in linkage map order have not been forthcoming. However, there has been a large number of chromosomal rearrangements detected in individual animals and, in some cases, their immediate relatives, using chromosome staining in pigs (Henricson and Bäckström, 1964; Popescu and Boscher, 1986; Ducos *et al.*, 2002), cattle (Christensen *et al.*, 1992; Pinton *et al.*, 1997; Joerg *et al.*, 2001) and chickens (Ramos *et al.*, 1999). The primary reason why chromosomal abnormalities are not detected in linkage studies is the associated reduction in fertility of carriers (Gustavsson, 1980). This can be mediated by pre-zygotic mechanisms such as an increase in non-disjunction of chromosomes or chromosome segments at meiosis that result in gametes with abnormal cytology and by a reduction in sperm fertility through decreased motility and abnormal morphology (Guttenbach

et al., 1997) or by post-zygotic mechanisms including an increase in spontaneous abortions of foetuses (Kalousek and Lau, 1992). During meiosis, selection against a chromosomal inversion can be reduced through the formation of a loop structure during the first metaphase (McClintock, 1931; McClintock, 1933) or the suppression of recombination in the region (e.g. Martin *et al.*, 1994; Brown *et al.*, 1998). Such suppression due to the presence of a small inversion has previously been demonstrated to be the cause of a fine scale difference in recombination rate between bulls (Park *et al.*, 1995; Park *et al.*, 1999). It follows that selection against an inversion may primarily occur at a post-zygotic level. This may be detected as the unequal transmission of alleles in the inverted segment of a heterozygous animal.

In the previous chapter, an inconsistency was detected between the published sheep linkage map (Maddox *et al.*, 2001) and the linkage map constructed from a commercial population of Charollais sheep. When compared to the published linkage map, markers MCM137 and BM6506 on chromosome 1 were in an inverted order relative to the flanking markers BMS527 and BM8246. In this chapter the region is examined with more detail in the Charollais population with the genotyping of two further markers. The region is also examined in an experimental population of Scottish Blackface sheep and a feral Soay sheep population.

3.2 Materials and Methods

Animals: Three sheep breeds were examined at the markers of interest. One of theses breeds, the Charollais, was studied through sire referencing schemes in commercial populations from the United Kingdom. The second, the Scottish Blackface, was an experimental sheep population at Roslin Institute. The third

breed examined was a feral Soay sheep population from St. Kilda, Scotland (see Chapter 5 for a description of this population). The Charollais sheep pedigree has been described in Chapter 2. Briefly, a complex pedigree of 570 animals derived from the descendants of five widely used sires was selected for a QTL mapping study. Of the 570 animals, 420, mainly descendants, were genotyped. The Scottish Blackface sheep pedigrees consisted of nine half-sib families of sizes ranging from 11 to 145, averaging 72 sheep. The Soay sheep pedigree was selected from a larger pedigree constructed using molecular techniques with the aim of maximizing power for QTL detection. Firstly, half-sibships of twelve or greater animals and their common parent were selected followed by the addition of half-sibships of at least ten animals that were linked to previously selected animals. This selected 588 animals for genotyping. The pedigree was extended further by including all available ancestral information for the selected individuals resulting in a total pedigree size of 868 animals.

Genotyping: The region of chromosome one showing a marker order inconsistency in the Charollais sheep population was further investigated with the genotyping of two additional microsatellite markers, BM7145 and BMS4008. The markers BM6506 and MCM137, which originally indicated the marker order inconsistency, were also re-genotyped to verify the accuracy of the commercial genotyping service. All four of these markers as well as two flanking markers from the Charollais sheep map, BM8246 and BMS527, were genotyped in the Scottish Blackface and Soay sheep pedigrees. Additional flanking markers (DB6, TGLA415 and SOX2) were genotyped in the Scottish Blackface pedigree to increase the precision of estimation of marker order as not all families were informative for the flanking markers. PCR reactions were performed with a 10μL reaction volume (Table 3.1) using a Touch-Down procedure which lowered the annealing temperature from 60°C to 50°C in 1°C steps during the first 11

Table 3.1: PCR reaction mixture protocol.

Component	Amount (μL)
ddH_2O	5.05
dNTP (2mM)	1.00
BioLab 10x NH_4 Reaction Buffer [a]	1.00
$MgCl_2$ (50mM)	0.50
F+R Primers (1μM) [b]	0.40
Taq-Polymerase (5u/μL)	0.05
Genomic DNA	2.00

[a] 160mM $(NH_4)_2SO_4$, 670mM Tris-Cl (pH 8.8 at 25°C), 0.1% Tween-20.

[b] Total primer aliquot, equivalent to 20μL of each primer.
See Table 3.2 for primer sequences.

cycles. The annealing temperature was then held constant for a further 19 cycles. Microsatellite allele lengths were determined with an ABI3730 DNA Analyzer. Genotypes were automatically determined using GeneMapper v3.0 (Applied Biosystems) followed by manual checking of the results.

Construction of Linkage Maps: Linkage maps were constructed using CRIMAP (Green *et al.*, 1990). CRI-MAP is widely used in the construction of linkage map in livestock populations due to its ability to handle large pedigrees from out-bred populations. This is achieved by the simplification of the full likelihood model through the exclusion of individuals who are uninformative or potentially uninformative for linkage (i.e. individuals that are homozygous at a locus or those that have a missing genotype that could possibly be homozygous). This simplifies the likelihood by avoiding the summation over all possible genotype probabilities in individuals with missing data. As the information ignored is consistent across all marker orders, the likelihoods of different orders are still able to be compared. As potential false 'double recombinant' individuals may result in a bias in linkage

Table 3.2: Forward (F) and reverse (R) primer sequences for genotyped markers.

Primer	Sequence
BM6506-F	GCACGTGGTAAAGAGATGGC
BM6506-R	AGCAACTTGAGCATGGCAC
BM7145-F	ATTATGTTCCAGATTCCATTCCA
BM7145-R	CAGCACTGTTTCATAAACTATGGG
BM8246-F	AATGACAAATTGAGGGAGACG
BM8246-R	AGAGCCCAGTATCAATTCTTCC
BMS4008-F	CGGCCCTAAGTGATATGTTG
BMS4008-R	GAAGAGTGTGAGGGAAAGACTG
BMS527	TCAGTGAAAGCAAGAGAAATATCC
BMS527	TCCATTCCCTTTGAATATCCC
MCM137-F	AGGGGAGCCCCAGTAAGTATCA
MCM137-R	AAACAAGTGGGGATGTTAGCTCTT

map construction, these need to be removed prior to linkage map construction. This is especially important in the central markers of the region studied as these cover only a few centiMorgans and are very unlikely to contain true double recombinants. However, the detection of such individuals requires prior knowledge of the linkage map order. This was accounted for using the following multi-stage procedure. Initially, the published linkage map order was assumed and potential double recombinant individuals were detected using the "chrompic" option of CRI-MAP and removed from the pedigree. On average this removed five individuals from the pedigree. Then the likelihood of all permutations of non-flanking marker orders was calculated. If a more likely marker order was detected, the removed individuals were replaced in the pedigree and the process was repeated using the new marker order. All map likelihoods in this study were calculated using a sex-averaged recombination distances, although this

reduced to the male linkage map in the case of the Scottish Blackface population where recombination information was available only from male individuals. In the other populations, specifying a linkage map for each sex provided the same results (data not shown).

The likelihood of all permutations of non-flanking marker orders was calculated. The five orders having with the largest log-likelihoods were recorded so that potential type-I error could be assessed. The current best estimate of the linkage map from the (AgResearch) International Mapping Flock (IMF) used in the construction of the published map was obtained from the framework given in Maddox *et al.* (2001) with distances as given on the Australian Sheep Genetics web site (Maddox, 2003). Only the central markers used in this study were placed at specific points on the published framework map but the potential positions shown for the flanking markers were both outside this region. This indicates the IMF population has no other marker order with likelihood within three LOD of the order considered here and has been confirmed with the original data (results not shown).

Detection of Transmission Distortion: As outlined above, post-zygotic selection against a chromosomal rearrangement can result in unequal transmission of chromosome types to a heterozygous individual's offspring. The detection of transmission distortion is achieved in two ways. The first approach involved an application of the Bradley-Terry model for ranking sports teams (Bradley and Terry, 1952). In this case, the alleles at a marker locus are ranked according to an indicator of transmission potential and tested for deviation from equality (Sinsheimer *et al.*, 2000). The analysis was performed using the "Gamete_competition" module of Mendel (v5.7; Lange *et al.*, 2001). Alleles with low frequencies in the population (>5%) were pooled to remove potential

bias associated with the use of large sample results in evaluating significance. As this single marker analysis potentially loses power due to not incorporating transmission information from multiple markers, an additional multipoint test for transmission distortion was also performed. The basis for the multipoint statistic is the observation that preferential transmission at a locus results in an increase in relatedness among an individual's offspring. Thus, an indicator of transmission distortion in a general pedigree of size n can be constructed by

$$TDI = \sum_{i=1}^{n} \sum_{j=1}^{i} \hat{\pi}_{ij} \tag{3.1}$$

where $\hat{\pi}_{ij}$ is the estimated identity-by-descent (IBD) probability for animals i and j. Multi-point IBD values were estimated by Markov-chain Monte Carlo (MCMC) sampling using LOKI (Heath, 1997) and assuming the most likely map order in each population. Average IBD coefficients were calculated from the mean of 10000 samples taken at every tenth iteration after a 1000 iteration burn-in. The empirical distribution of the indicator statistic was assessed by one million gene-dropping simulations. In these simulations, all founders were assumed to have unique alleles as this allowed for rapid calculation of IBD values.

3.3 Results

Repeated Genotyping: Comparisons of allele lengths obtained by the commercial genotyping service and the repeated genotyping for markers MCM137 and BM6506 in the Charollais sheep pedigree are shown in Figures 3.1 and 3.2 respectively. Marker MCM137 shows a good concordance between the commercial and repeated genotyping. However, BM6506 shows a consistent difference in a approximately 10% of the genotypes. As the repeated genotypes for both MCM137 and BM6506 were analysed on the same plate and subsequent genotype determination was performed on all animals at once, the source of the

Table 3.3: Summary of marker information in the three examined populations.

Population	Marker	Number of Alleles	PIC[a]	Informative Meioses
Charollais	BMS527	8	0.69	373
	MCM137	15	0.83	494
	BM7145	7	0.72	332
	BM6506	6	0.70	308
	BMS4008	7	0.70	366
	BM8246	9	0.82	483
Soay	BMS527	5	0.42	153
	MCM137	7	0.70	176
	BM7145	4	0.64	112
	BM6506	5	0.63	153
	BMS4008	5	0.74	126
	BM8246	5	0.59	128
Scottish	BMS527	7	0.75	328
Blackface	*DB6*	16	0.78	378
	MCM137	15	0.89	383
	BM7145	8	0.60	189
	BM6506	6	0.63	343
	BMS4008	9	0.73	356
	TGLA415	10	0.52	191
	SOX2	17	0.73	375
	BM8246	8	0.60	37

[a] PIC - Polymorphic Information Content (Botstein *et al.*, 1980).

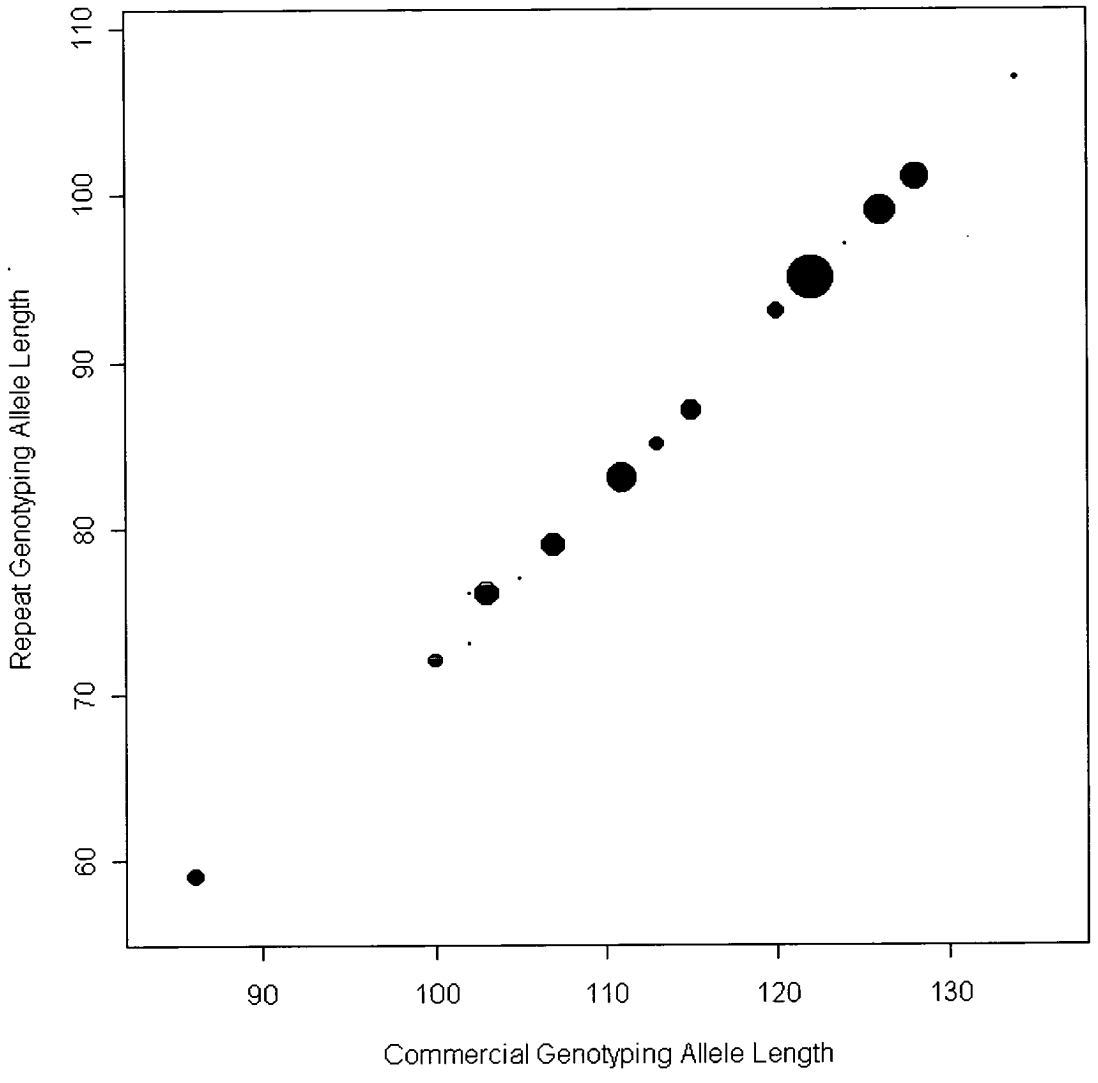

Figure 3.1: Comparison of microsatellite allele lengths for MCM137 obtained by the commercial genotyping service and the repeated genotyping. A good concordance between results is shown with only one genotype deviating from a 1-to-1 relationship between replicates. Allele length differed by 28 or 29 bases between replicates.

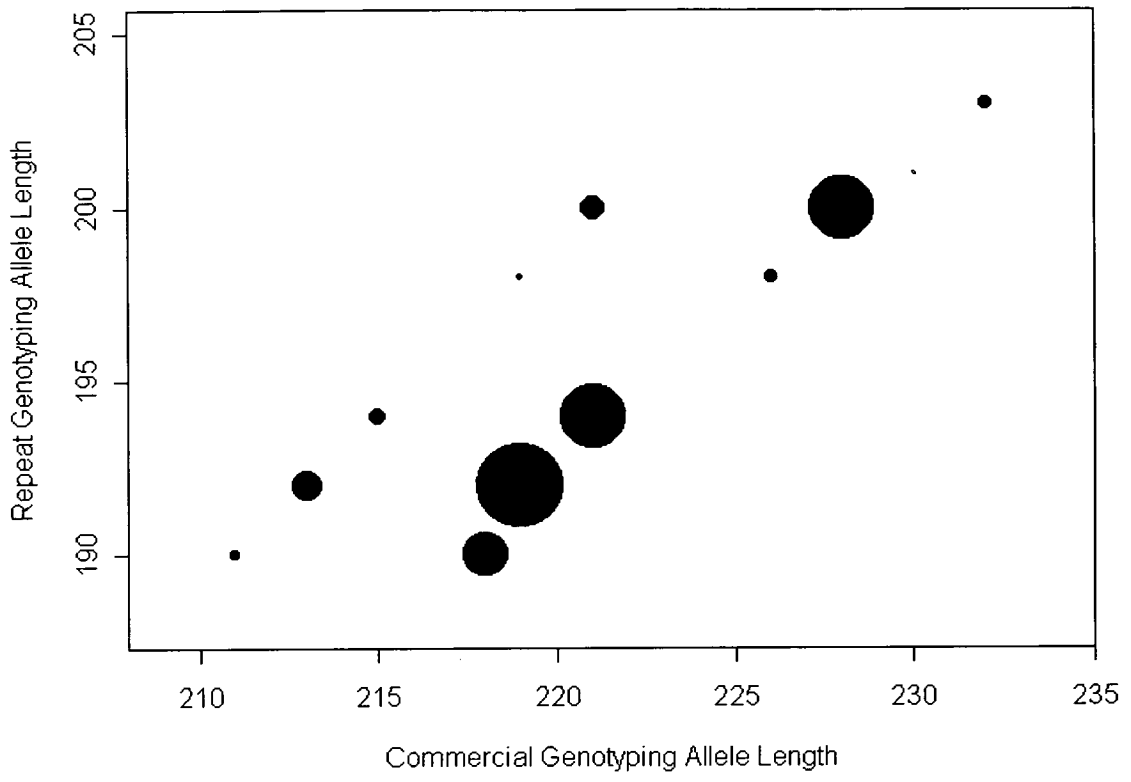

Figure 3.2: Comparison of microsatellite allele lengths for BM6506 obtained by the commercial genotyping service and the repeated genotyping. A lack of concordance between replicates is evident by the presence of two parallel lines. The nature of the difference indicates a consistant methodological error has occurred in a small group of the animals genotyped. The differences in allele lengths between replicates is 27-29 bases for the larger group and 21 bases for the smaller group.

error appears to be the commercial genotyping company. Thus, the repeated genotyping allele lengths were used in the analysis that follows.

Linkage Map Construction: The estimated linkage maps for the populations for the region examined are given in Table 3.4. In the Charollais and Soay sheep populations, the most likely map order was greater than three LOD more likely than any other marker order. These maps are graphically represented in Figure 3.3. A summary of the five most likely map orders in the three populations

Table 3.4: Estimated marker positions for the most likely linkage map orders in each population. No other marker orders were within three LOD of the mostly likely order. The relative position of each marker is given beside the estimates.

Marker	Population			
	IMF[a]	Charollais	Scottish Blackface	Soay
BMS527	0.0 (1)	0.0 (1)	0.0 (1)	0.0 (1)
MCM137	18.7 (2)	18.0 (3)	23.0 (3)	19.6 (2/3)
BM7145	19.8 (3)	18.6 (4)	22.6 (2)	19.6 (2/3)
BM6506	20.9 (4)	15.9 (2)	25.1 (4)	20.3 (4)
BMS4008	22.0 (5)	19.8 (5)	26.4 (5)	22.3 (5)
BM8246	27.2 (6)	25.6 (6)	34.2 (6)	29.7 (6)

[a] Marker distance is measured from BMS527 which is position at 213.9cM on chromosome 1 in the published linkage map.

examined is given in Table 3.5. The order of markers BMS527 and MCM137 was not able to be elucidated in the Scottish Blackface population with the most likely map order positioning these markers together. The Charollais population shows strong evidence against having the marker order given by the IMF populations (-10.0 LOD) and the order given by the Soay sheep population (-12.0 LOD). Similarly, the Soay sheep population provides evidence against having the IMF marker order (-3.2 LOD) and the order in the Charollais population (-12.8 LOD). The marker order obtained in the Scottish Blackface populations is consistent with the orders given in both the IMF and Soay populations but not the Charollais order (-3.2 LOD).

Detection of Transmission Distortion: The significances of the single-point analysis of allele transmission are given in Table 3.6. Only the Charollais population shows

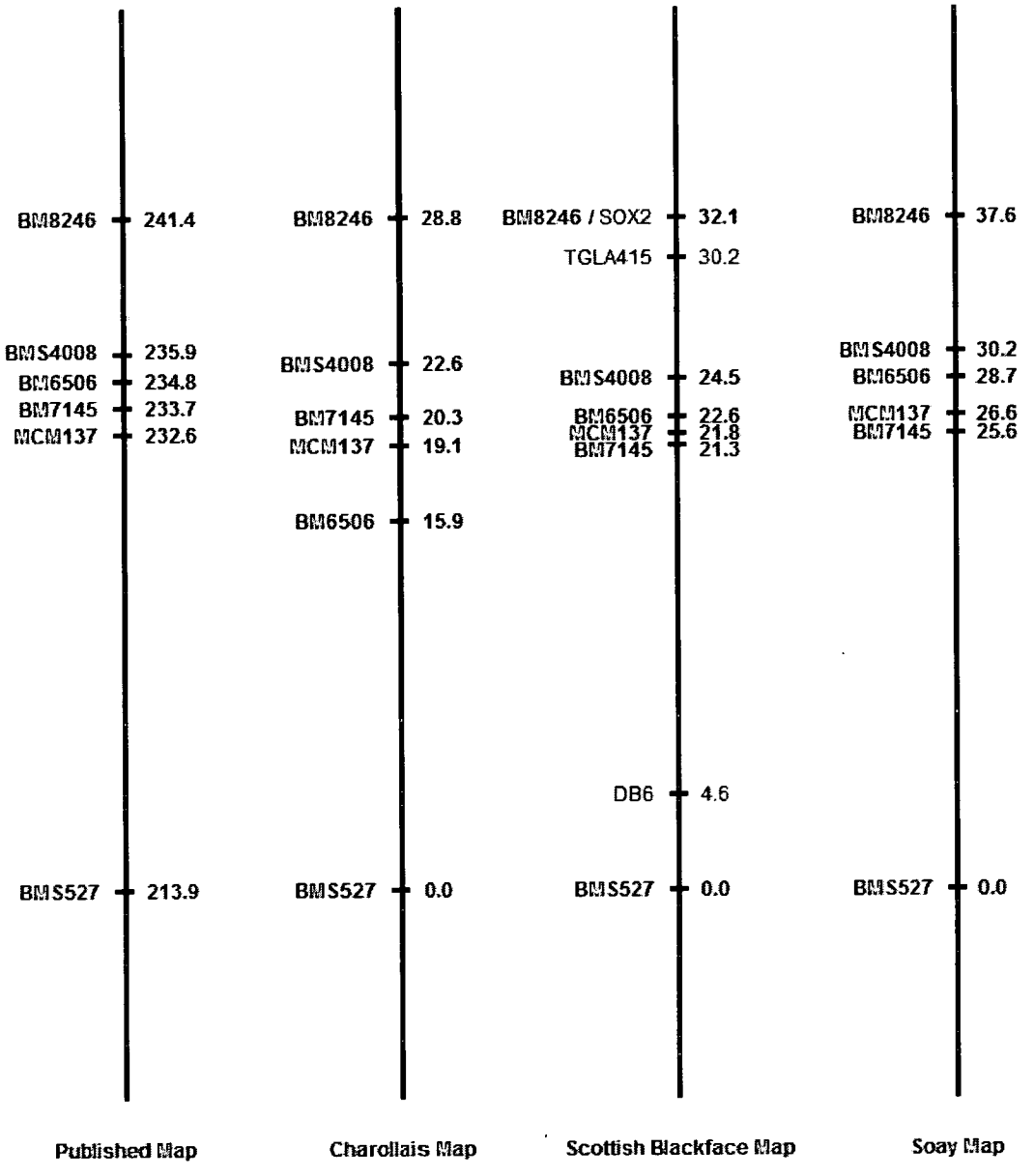

Figure 3.3: Graphical representation of the most likely map orders in each population. The order of marker BMS4008 relative to the flanking markers BMS527 and BM8246 remains unchanged in all populations, as does BM6506 in all populations except the Charollais sheep. The position of one of the extra markers, SOX2, in the Scottish Blackface population could not be distinguished from that of BM8246 but was set to be between BM8246 and TGLA415 in the analysis as given in the published map.

Table 3.5: Summary of the five most likely marker orders for each population. Locus numbers represent the order from the IMF population starting at BMS527. Markers with tied positions are indicated with a slash.

Population	Order	Log_{10}-Likelihood	Map Length
IMF	1,2,3,4,5,6	-	27.2
Charollais	1,4,2,3,5,6	-125.97	25.6
	1,4,2,5,3,6	-129.12	26.2
	1,4,3,2,5,6	-129.65	26.0
	1,4,5,2,3,6	-129.79	26.8
	1,4,5,3,2,6	-131.72	26.6
Soay	1,3,2,4,5,6	-124.48	34.2
	1,2,3,4,5,6	-127.63	34.8
	1,3,2,5,4,6	-130.05	35.0
	1,5,4,2,3,6	-132.99	40.4
	1,2,3,5,4,6	-133.22	35.6
Scottish	1,2/3,4,5,6	-168.51	29.7
Blackface	1,4,2/3,5,6	-171.64	31.1
	1,2,4,3,5,6	-172.59	30.2
	1,3,4,2,5,6	-173.98	30.9
	1,5,2/3,4,6	-180.21	32.8

any evidence of transmission distortion with marker BM7145 being significant at the 5% level. While this result could potentially be caused by problems with the PCR, for example uneven amplification of alleles, such problem were not apparent during the scoring of genotypes and would be expected to occur in all breeds examined. Also, there was no excess in Mendelian inconsistencies at this marker relative to other markers. The multi-point transmission distortion indicator statistic provides no significant results at any of the marker positions (Table 3.7). However, in the Charollais population the two flanking markers show less

Table 3.6: Significance of single-point analysis of transmission distortion. Only marker BM7145 in the Charollais sheep population shows significant amount of transmission distortion at the 5% level.

	Population		
Marker	**Charollais**	**Scottish Blackface**	**Soay**
BMS527	0.924	0.146	0.469
MCM137	0.755	0.361	0.518
BM7145	**0.015**	0.453	0.625
BM6506	0.416	0.157	0.323
BMS4008	0.150	0.899	0.708
BM8246	0.562	0.968	0.432

Table 3.7: Multi-point analysis of transmission distortion. The proportion of gene-drop replicates having larger transmission distortion indicator values than that calculated from marker data is shown. Although not significant, the central markers in the Charollais sheep linkage map show increased allele sharing compared to the flanking markers.

Marker	**Population**		
Position	**Charollais**	**Scottish Blackface**	**Soay**
1 (BMS527)	0.200	0.050	0.416
2	0.123 (BM6506)	0.237 (BM7145)	0.638 (MCM137)
3	0.132 (MCM137)	0.240 (MCM137)	0.639 (BM7145)
4	0.095 (BM7145)	0.314 (BM6506)	0.699 (BM6506)
5 (BMS4008)	0.078	0.447	0.671
6 (BM8246)	0.176	0.264	0.678

allele-sharing than all the internal markers, a situation expected if transmission distortion is occurring due to selection against chromosomal heterogeneity in this region. Marker BMS527 in the Scottish Blackface population approaches significance in the multi-point analysis but not in the single-point analysis. However, it provides no evidence for selection on a chromosomal rearrangement as it is a flanking marker for the region of interest.

3.4 Discussion

With the increasing effort to fine-map and ultimately locate variants in regions detected to have an effect on commercially important traits in previous genome-wide scans, it is becoming increasingly important to have dense and accurate linkage maps. Here, a difference between the linkage maps of the IMF population and the Charollais and Soay sheep populations has been demonstrated. The map orders from the Charollais and Soay populations also differ from each other. The estimate marker order in the Scottish Blackface is consistent with both the order given by both Soay and IMF populations. Thus, this study has identified a total of three different map orders across four sheep populations.

The marker orders observed in this study require at least two chromosomal rearrangements since the sheep breeds last shared a common ancestor. The most likely order in the Soay sheep population shows a two marker inversion (MCM137 and BM7145) compared to the IMF linkage map. The relationship between the linkage maps for the Charollais sheep population and the other populations depends on which map is chosen as the correct map. The most likely order in the Charollais population is a three marker inversion (BM7145, MCM137, BM6506) from the order given in the Soay sheep population and requires either two inversions or a translocation to achieve the order given by the IMF population. This situation is reversed if the second most likely order is assumed for the

Charollais population. This suggests the most parsimonious solution to the ancestry of these sheep breeds is to consider the Soay sheep as an outgroup to the IMF and Charollais populations with a different inversion occurring in both the IMF and Charollais lines once these had split. This interpretation is consistent with the primitive status of the Soay sheep. Unfortunately, not all these markers amplify in other livestock species or, if they do amplify, their order cannot be elucidated using the current linkage mapping populations, so it is not possible to further clarify the history of this region.

This study provides some evidence for selection occurring in the region of interest in the Charollais sheep population through the observation of unequal transmission of alleles at BM7145 and an increase in allele-sharing at the central marker positions compared with the flanking marker positions. Both methods used to detect possible transmission distortion have potential pitfalls that need to be examined. The single-point analysis ranks the alleles at a marker locus based on their transmission probabilities when in competition with each other (i.e. in a heterozygous individual). Thus it is a test of association between marker alleles and the locus causing the transmission distortion. While it is likely that a chromosomal rearrangement occurs initially in one individual, any association with marker alleles is likely to be incomplete, thus reducing the power for detecting transmission distortion. Also, the requirement of the pooling rare alleles for the large sample distribution of the test statistic to be applicable may potentially combine alleles at the opposite end of the transmission spectrum, biasing the results towards the null hypothesis.

The multi-point test statistic has an underlying bias induced by the methodology used to estimate the null distribution. In the gene-dropping simulations, all founder individuals were considered to have unique alleles as this allowed for

rapid computation of IBD values that are otherwise computationally demanding in the complex pedigrees used. A consequence of this assumption is an increased range in the simulated null distribution. This can be most easily conceptualized by firstly considering the case where no marker information is available. In this case, the estimated IBD value for any pair of individuals is equal to the expected value for their given relationship and the variance is zero. As marker information is added, the expected IBD value remains the same but its variance increases. This results in an increased range for the transmission distortion indicator statistic with increasing marker information. The increase in range occurs primarily above the expected value due to the correlation of IBD value across related individuals. In fact, the minimum value of the transmission distortion indicator statistic in a group where any individual is equally related to all other individuals tends to the value calculated using expected IBD values as sample size increases. Thus the evidence for selection in the examined region is increased by the knowledge that marker inheritance is not fully determinable given the number of alleles at the marker loci.

The observation of heterogeneity of chromosome structure has implication for the design and analysis of fine-mapping studies. As the map cannot be assumed to be the same as the published map, information on linkage will also need to be collected. Thus, fine-mapping using linkage disequilibrium in a random sample from a population should be undertaken with caution. Instead designs such as several large half-sib families are more appropriate as all individuals provide information through both linkage and linkage disequilibrium. The choice of analysis methodology used in fine-mapping studies also depends on the amount of information about linkage that is available. While single locus methods, such as regression of trait values on genotype, do not directly require linkage information for their use, the following attempts to characterize the underlying

genetic variants in the detected region may be hampered without this information. Also, given the large amount of variability in the amount of linkage disequilibrium between pairs of loci separated by a fixed distance in livestock populations (Farnir *et al.*, 2000; McRae *et al.*, 2002; Nsengimana *et al.*, 2004), multi-locus methods that average out this variability are likely to be more powerful than single locus methods. All multi-locus methods will require a genetic map to be accurately known so the appropriate weighting of information at each locus is achieved. Incorrect specification of linkage maps may result the localization of a locus to an incorrect region.

3.5 Appendix: Transmission Distortion Through Post Zygotic Selection Against a Chromosomal Inversion

Consider an population segregating for a chromosomal inversion that is selected against *in utero*. Let the normal and inverted chromosomes be represented as C and C^* respectively and the frequency of C in the population be p where $0 < p < 1$. In addition, let the fitness of animals with genotypes CC^* and C^*C^* relative to those with genotype CC be $1 - s$ and $1 - t$ respectively, where $0 \leq s, t \leq 1$. An animal that is heterozygous for the inversion is expected to pass on chromosomes C and C^* at at the inverted segment with ratio of

$$1 - s + ps : 1 - t - ps + pt. \tag{3.2}$$

Thus, the ratio of transmission at the inverted chromosome segment is different from $1 : 1$ for general s and t unless $p = (s - t)/(2s - t)$ (see Figure 3.4). In the special case of $t = 0$ it is clear the equilibrium only exists at $p = 0.5$ for all values of s. All these points of equal transmission are unstable with genetic drift in transmission from heterozygotes and, when t does not equal zero, selection against C^*C^* homozygotes shifting the allele frequency away from that required

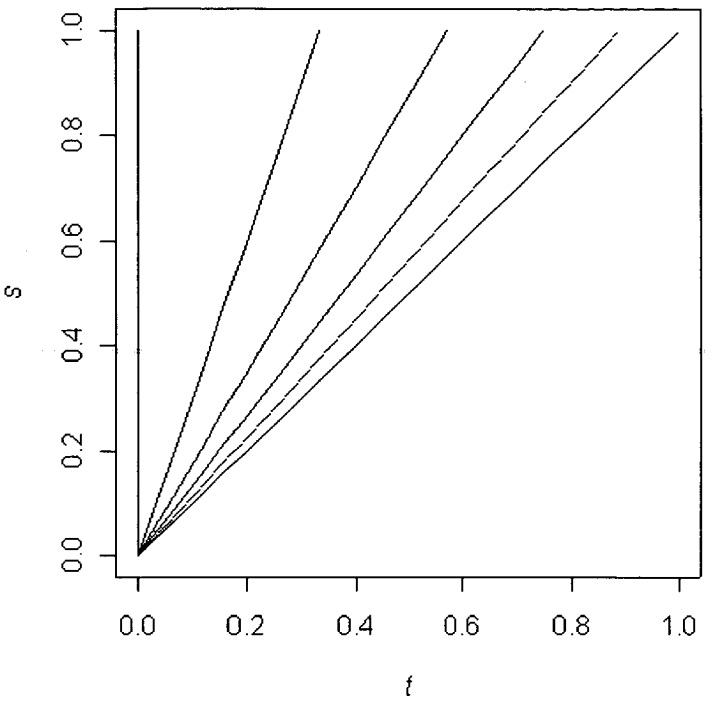

Figure 3.4: The effect of different parameter values on transmission distortion from a heterozygote for a chromosomal inversion that is selected against *in utero*. Each line represents the values of the selection coefficients, s and t, that result in a 1 : 1 transmission of chromosomal types with values of p, in order of increasing line slope, of 0, 0.1, 0.2, 0.3, 0.4 and 0.5. When pairs of selection parameter values fall in the region below the each line, C is transmitted preferentially to C^* with the opposite transmission above the line. For $p > 0.5$ all non-zero selection values result in preferential transmission of C. Note the line for $p = 0$ is an asymptotic result.

for equal transmission. Thus the only stable equilibrium points are $p = 0$ and $p = 1$ where there is no variation for selection to act upon.

4 Evaluation of Significance Thresholds for QTL Mapping in Complex Pedigrees by Partial Trait Permutation

4.1 Introduction

The localization of quantitative trait loci (QTL) to regions of the genome using multiple linked genetic markers requires the determination of an appropriate significance threshold the test statistics can be compared against. This threshold is required to account for the large number of test performed at intervals across the genome, many of which are not independent.

A number of approaches have been developed for the estimation of significance thresholds. The most direct approach is to assess the threshold theoretically. Under the assumptions of infinitely dense and fully informative markers and recombination events being independent (i.e. occurring without interference), thresholds have been derived for QTL mapping in experimental populations (Lander and Botstein, 1989; Kruglyak and Lander, 1995; Dupuis and Siegmund, 1999; Zou *et al.*, 2001). In some cases, the requirement of infinitely dense markers can be reduced to a finite set of equally spaced markers provided that tests are only performed at marker positions (Dupuis and Siegmund, 1999; Zou *et al.*, 2001). These results form the basis of the widely used guidelines for assessing the significance of a linkage statistic provided by Lander and Kruglyak (1995). Parametric simulation studies provide a useful alternative in cases where the theoretical evaluation of the distribution of the linkage test statistic under the null hypothesis is not tractable. However, these rapidly become computationally expensive as the complexity of the data increases. Another disadvantage of these methods is that the significance thresholds are highly dependant on the parametric assumptions used in their construction.

A second class of approaches attempts to estimate the number of independent tests performed directly from the data. With this information a simple Bonferroni correction can be performed to obtain an appropriate significance level. Cheverud (2001) proposed achieving this by using the variance of the eigenvalues of a correlation matrix of marker data. A higher correlation between marker data leads to an increase in the variance of the eigenvalue matrix. This is most easily visualized in the extreme cases of completely dependent or independent marker data. When marker data are independent, and thus show zero correlation, the eigenvalues of the correlation matrix will all be equal to one and thus have zero variance. With fully dependent marker data, one eigenvalue of the correlation matrix will equal the number of markers and all others are zero, giving a large variance for the eigenvalues. This approach is most readily applied in structured pedigrees where marker data are used in a linear regression framework such as Haley-Knott regression in F_2, back-cross or half-sib families (Haley and Knott, 1992; Haley *et al.*, 1994; Knott *et al.*, 1996). However, it is not clear how the correlations should be calculated for the use of this method in more complex pedigree structures.

A more general approach involves the detection of the correlation structure in the data indirectly through the path of the linkage test statistic across the genome. These are based on the work of Davies (1977; 1987) who suggested estimating a significance threshold using the points of inflection of correlated test statistics (i.e. points where the test statistic reaches a local maxima or minima). Rebaï *et al.* (1994; 1995) use this method to derive an upper bound of the significance threshold for the cases of a backcross and F_2 population. The application of this method to more general pedigree structures has been demonstrated by Piepho (2001). However, this method is limited to the case where a test

statistic follows a chi-squared distribution and thus cannot be applied to variance component linkage mapping where the test statistic follows a mixture of zero and chi-squared under the null hypothesis (Self and Liang, 1987). Recently, a very similar approach that models the test statistic as an auto-regressive process has been proposed (Bacanu, 2005). One disadvantage of these approaches is that when only one or a few chromosomes are examined, the thresholds will be estimated with a large variance.

Another method for estimating significance levels involve the exploration of the parameter space of the test statistic under the null hypothesis through the use of permutation (Churchill and Doerge, 1994). Any phenotype - genotype correlation is removed by permuting the original phenotypic data over the fixed genetic data. Those permuted data are then analysed providing a random deviate from the distribution of the test statistic under the null hypothesis of no QTL. This is repeated until the distribution of the test statistic under the null hypothesis is known with enough accuracy to estimate the required significance threshold. Permutation methods have the advantages of being distribution free and thus they do not depend on distributional assumptions that may not be satisfied by the data being analysed (Doerge and Rebaï, 1996). While initially proposed for the use with simple experimental designs, permutation methodology may also be used with simple structured data sets such as sib-pairs (Wan *et al.*, 1997). As the computation time required to give precise significance levels may be large, approaches that test significance at a pre-defined level have been developed (Nettleton and Doerge, 2000).

While it may not be immediately obvious, permutation methods are not able to be directly used with a variance component analysis of general pedigrees. Under

the null hypothesis, the data are modelled as

$$\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e} \tag{4.1}$$

where $\mathbf{y}$ is vector of phenotypes, μ is the mean of the phenotypes, $\mathbf{Z}$ is an matrix relating individuals to phenotypes, $\mathbf{u}$ is an vector of additive polygenic effects and $\mathbf{e}$ is the residual vector. It follows that the null hypothesis is more complicated than just the lack of a QTL at the point of the genome being tested, it also includes information about the proportion of the variation in the trait attributable to genetic components. Thus, the null hypothesis is better stated as "none of the (additive) genetic contribution of the trait is attributable to this point in the genome". It now becomes clear that permuting under the null hypothesis must maintain the information contained in the pedigree structure. With simple structured pedigrees, all individuals are equally related and thus permutation does not alter any information given in the pedigree. This is not the case with more complex pedigrees.

In Chapter 2, a variance component linkage analysis was performed on a complex sheep pedigree. As none of the methods described above is directly applicable to such studies, the threshold for significance was obtained from published simulation studies (Xu and Atchley, 1995; Grignola *et al.*, 1996). Both these studies suggested a threshold somewhere between those for a χ^2_1 and a χ^2_2 distribution. However, while these studies simulated very different pedigree structures (500 full-sib families with two offspring in Xu and Atchley and a granddaughter design with 2000 sons and 20 sires of sons in Grignola *et al.*), the genetic and trait data were very similar consisting of five or six markers separated by 20cM and a trait heritability of 0.5. Thus, the suggested thresholds should be used with caution in cases where the data structure differs from those simulated.

An ideal solution to this problem would be the extension of one of the above methods for calculating significance thresholds to complex pedigree structures. In this chapter, the application of permutation methodology to pedigrees with increased complexity is examined.

4.2 Materials and Methods

Permutation of Trait Values in General Pedigrees: The objective of the permutation of trait values in the evaluation of significance thresholds is to randomize the trait values with respect to identity-by-descent measures while maintaining the underlying structure of the data. In non-experimental systems, this structure needs to include the familial relationships between individuals so that the estimated heritability of the analysed trait remains constant. The trait values for individuals i and j can be permuted without altering the observed trait heritability when $\theta_{ik} = \theta_{jk}$ for all $k \neq i, j$, where θ_{xy} is the coefficient of coancestry or the expected additive genetic relationship between individuals x and y. It can be shown that if individuals A and B and individuals A and C satisfy this permutation condition, then individuals B and C also satisfy the condition. Thus, it is possible to form independent groups of individuals within which trait values are able to be permuted. The resulting permutation test is referred to as a partial permutation as, in general, this methodology will result in only a part of a pedigree being permuted (see Figure 4.1). An important property of this partial permutation method is that it reduces to the standard permutation methodology when used with designs using structured pedigrees (such as a backcross or F_2 design).

Efficiency of Partial Permutation: The partial permutation methodology will not randomize all the information about linkage in a general pedigree. For example, in Figure 4.1 the sib-pair in the second generation, and through

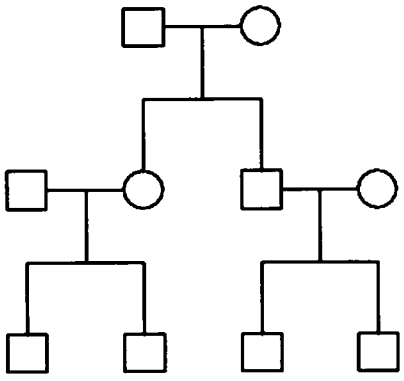

Figure 4.1: Example of a simple three generation pedigree. Under the restrictions of partial permutation, the two individuals in the first generation and the sib-pairs in the third generation are permutable.

them their offspring, will provide linkage information that is not removed by partial permutation. A measure of the proportion of a pedigree being permuted is needed in order to assess the efficiency of the partial permutation in reconstructing the null distribution. Initially, taking the ratio of the number of pairs of individuals permuted to the total possible pairs may appear to be an appropriate measure as this returns a value of zero when no permutation is performed and one when all individuals are permutable. However, this does not take into account the variation in the amount of information supplied to the linkage test statistic across pairs of individuals. For example, in the absence of inbreeding, parent-offspring pairs supply no information about linkage and thus not permuting these pairs will not cause a reduction in permutation efficiency. Also, there is no accounting for the possibility that while a pair may not be directly permutable, the permuting with other individuals may result in a proportion of the information supplied by this pair being randomized.

An improved measure of permutation efficiency can be constructed by considering any permutation to involve the swapping of the relevant values of the estimated identity-by-descent matrix rather than trait values. Formally, if individuals i

and j are permutable, then permuting their trait values is equivalent to setting $\hat{\pi}_{p,ik} = \hat{\pi}_{jk}$ and $\hat{\pi}_{p,ki} = \hat{\pi}_{kj}$ for all individuals $k \neq i, j$, where $\hat{\pi}_{ij}$ is the estimated IBD value for individuals i and j and $\hat{\pi}_{p,ij}$ is their permuted IBD value. The change in the IBD matrix due to the permutation can be measured as

$$\delta_p = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\hat{\pi}_{ij} - \hat{\pi}_{p,ij})^2 \tag{4.2}$$

where n is the number of individuals in the pedigree. This measure accounts for the variation in the amount of information provided by different pairs of individuals in the pedigree as the summand is related to the variance of possible IBD values for individuals i and j which in turn is related to the amount of information contributed by this pair to the linkage test statistic (Williams and Blangero, 1999). Also, pairs of individuals whose trait values are not permutable may still increase the permutation efficiency if any permutation involves their IBD value.

For practical application, several aspects of this statistic need to be considered. Firstly, as the value of δ_p varies with the actual permutation, an average across permutation replicates, $\overline{\delta}_p$, will need to be used. Another desirable property would be for this measure to be comparable across pedigrees. This can be achieved by scaling $\overline{\delta}_p$ by its value in the ideal situation where permutation effectively provides an IBD matrix from an independent observation of genetic inheritance on the same pedigree. This ideal value will be referred to as $\overline{\Delta}_p$. A final consideration in the practical implementation of this statistic is the variation in marker informativeness across the genome. As the statistic is based on IBD values, its value will vary with the amount of marker information. This can be accounted for by calculating the statistic using gene-dropping simulation where all founder alleles are unique. Thus, the following steps are

used in calculating the efficiency of partial permutation. Firstly, a random inheritance vector is obtained for the pedigree using gene-dropping methodology with all founders' alleles being unique. The calculation of IBD values is made trivial in this situation. This IBD matrix can be permuted and a value of δ_p is calculated. Repeating the gene-dropping step provides a second IBD matrix that is compared to the first giving a value for Δ_p. These steps are repeated until the efficiency statistic, $\overline{\delta}_p/\overline{\Delta}_p$, is estimated at the desired accuracy.

Analysis of Simulated Datasets: The performance of the estimation of significance through the use of partial permutation was assessed through the use of simulation studies on two pedigree (Figures 4.2 and 4.3). These pedigrees represent multiple F_2 families sharing the same P_1 grand-parents but different F_1 parents. Such a design might be employed when the size of families is constrained or if there was potentially genetic variants for the trait of interest segregating within the parental lines. Both example pedigrees contain 122 individuals but differ in the number and size of the F_2 families. The first pedigree (Figure 4.2) contains 12 F_2 families of size eight and the second pedigree (Figure 4.3) has 30 families of size two. These pedigrees will be referred to as 'Pedigree A' and 'Pedigree B', respectively.

Genetic data were simulated as a chromosome of 100cM in length containing six equally spaced markers. Recombination rates between adjacent markers were calculated using Haldane's map function. The founding four chromosomes were given unique alleles in order to allow the rapid calculation of IBD matrices. Trait values were simulated under a purely additive model with a heritability, h^2, of 0.5. In models with no QTL effect, additive genetic effects of founder individuals were drawn from a normal distribution with variance $a^2 = h^2$. The additive genetic effect in descendants was calculated as the average of the effects

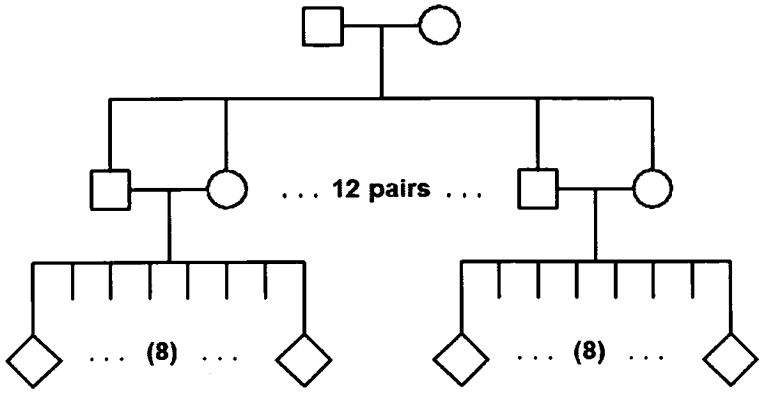

Figure 4.2: Simulation Pedigree A. The 122 individuals are structured as twelve F_2 families with eight offspring descended from the same P_1 grandparents.

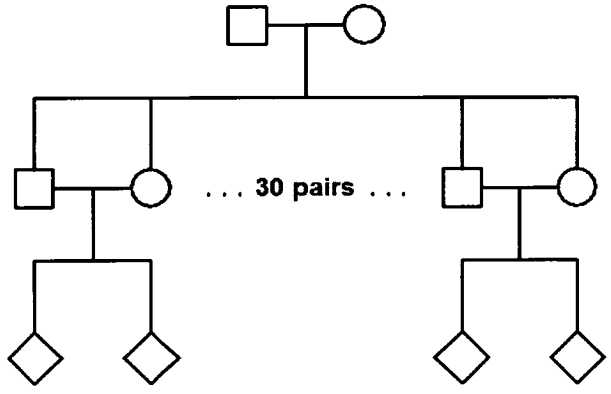

Figure 4.3: Simulation Pedigree B. The 122 individuals are structured as 30 F_2 families with two offspring descended from the same P_1 grandparents.

in their parents plus a random normal Mendelian sampling effect with variance of $a^2/2$. In models including a QTL effect, a QTL locus was positioned at 50cM along the chromosome. QTL allelic effects for the founder chromosomes were drawn from a normal distribution with variance $q^2/2$ and were inherited in the usual Mendelian fashion. The residual additive genetic effect was simulated as the additive genetic effect in the case with no QTL, but with a variance of $b^2 = a^2 - q^2$. Specific environmental effects for all individuals were drawn from a normal distribution with variance $e^2 = 1 - a^2$.

The simulated data were analysed using a variance component framework. This partitioned the variance in the trait data into additive genetic and environmental components under the null hypothesis and further partitioned the additive genetic effect into components due to the chromosome position being analysed and the remainder of the genome under the alternative hypothesis. A more detailed description of the variance component models is given in Chapters 1 and 2. Parameter estimates were calculated using restricted maximum likelihood (REML) with the program ASREML (Gilmour *et al.*, 2002). To reduce the time of computation, models were only fitted at the six marker positions along the chromosome.

Thresholds for chromosome-wide significance at the 5% level were estimated from the 95% quantile of the distribution of the maximum test statistic across the chromosome of 10000 simulated data sets with no QTL effect. Confidence intervals for the estimated threshold were calculated using standard bootstrap methodology (Efron, 1979; Efron, 1981). This involved estimating the distribution of the sampling variation in the estimated threshold using 1000 bootstrap samples from the simulated null distribution. A symmetric 95% confidence interval can then by calculated by taking the appropriate quantiles of the bootstrap distribution. The distribution of partial permutation thresholds was estimated from 1000 simulation replicates each performing 1000 permutation analyses. Confidence intervals for the mean permutation threshold were calculated using bootstrap methodology analogous to that described above. The effect of the presence of a QTL on the permutation thresholds was investigated by simulating QTL effects of size $q^2 = 0, 0.1$ and 0.25.

Table 4.1: Estimated significance thresholds and approximate 95% confidence intervals.

Method	Threshold	
	Pedigree A	Pedigree B
Simulation	4.77 (4.56, 4.88)	4.46 (4.28, 4.62)
Permutation		
$q^2 = 0$	4.41 (4.34, 4.48)	3.88 (3.78, 3.98)
$q^2 = 0.1$	4.87 (4.79, 4.94)	5.53 (5.37, 5.72)
$q^2 = 0.25$	5.59 (5.48, 5.71)	8.46 (8.18, 8.76)

4.3 Results

The efficiency of partial permutation in the two simulated pedigrees was estimated using the average value over 10000 gene-dropping replicates giving efficiencies of 80% and 63% in Pedigree A and Pedigree B respectively. The average significance thresholds obtained from the simulated data and their 95% bootstrap confidence intervals are given in Table 4.1. The empirical distributions of the thresholds obtained using partial permutation are given in Figure 4.4. For both pedigrees, the partial permutation threshold estimated in the case of no QTL is significantly lower than the threshold calculated directly through simulation. As the size of the QTL effect increases, the threshold estimated through using partial permutation increases with the thresholds for the case with QTL effect of 0.25 being significantly higher than calculated directly through simulation. The size of the bias in threshold values is greater for Pedigree B where fewer individuals are able to be permuted.

The observed upward bias in thresholds for cases with non-zero QTL effect sizes is readily explained as some of the information about QTL linkage will

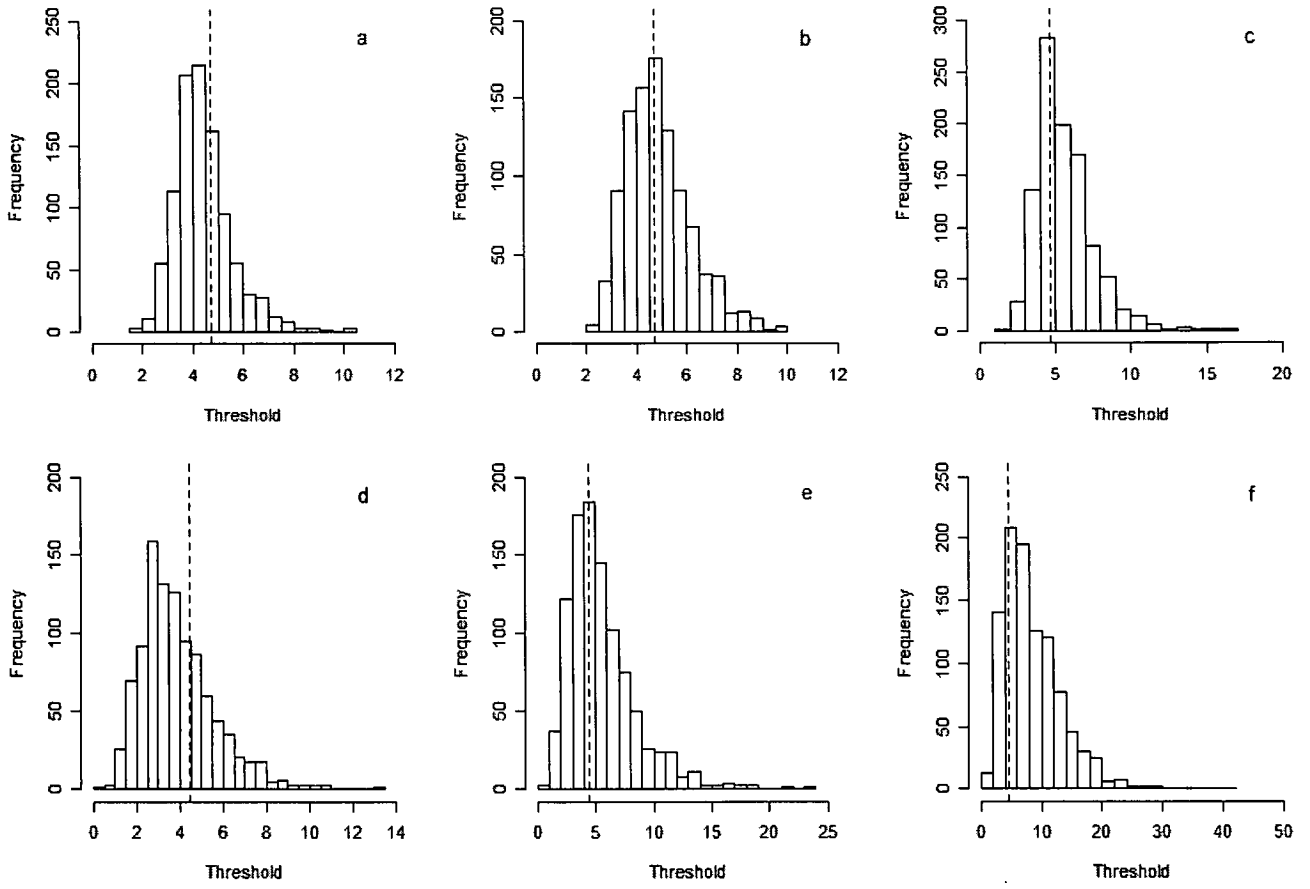

Figure 4.4: Distribution of 5% significance thresholds obtained under the various sampling conditions. Plots a, b and c are for Pedigree A with QTL effects of $0, 0.1$ and 0.25 respectively. Plots d, e, and f are the corresponding distributions obtained using Pedigree B.

remain in the partially permuted data. However, the downward bias observed when using partial permutation in the case with no QTL warrants further investigation. When the distributions of the maximum test statistics obtained by partial permutation in all simulation replicates are pooled, this distribution is nearly identical to that obtained by direct simulation. For both pedigrees examined, the mean and standard deviations of these two distributions are different only from the second decimal place. However, when replicate simulations are considered independently, the distribution of test statistics obtained by partial permutation varies markedly from this average. The mean and variance of the distributions of the maximum test statistic across the chromosome obtained by partial permutation in Pedigree A are given in Figure 4.5. The strong correlation between these means and variances can be explained by considering the contribution to the test statistic from the proportion of the pedigree which is not permuted. Under the null hypothesis, the test statistic for linkage in the entire pedigree is a random deviate from its null distribution. If this value is small, then the test statistic calculated by considering just the non-permutable portion of the pedigree will on average be from the lower end of its null distribution. Similarly, if the overall test statistic is large, that from the non-permutable portion will also be larger on average. The size of the test statistic for the non-permutable portion of the pedigree then determines the possible range of the permutation test statistics on the whole pedigree. For example, with a small test statistic for the non-permutable part of the pedigree, the possible range of test statistics for the permuted pedigree is reduced compared to the exact null distribution as the upper portion of possible values will not be achievable. It follows that the upper tail of the distribution of the test statistic is being explored by partial permutation in only a few simulation replicates and thus unbiased estimates of the upper quantiles are not achievable using this method.

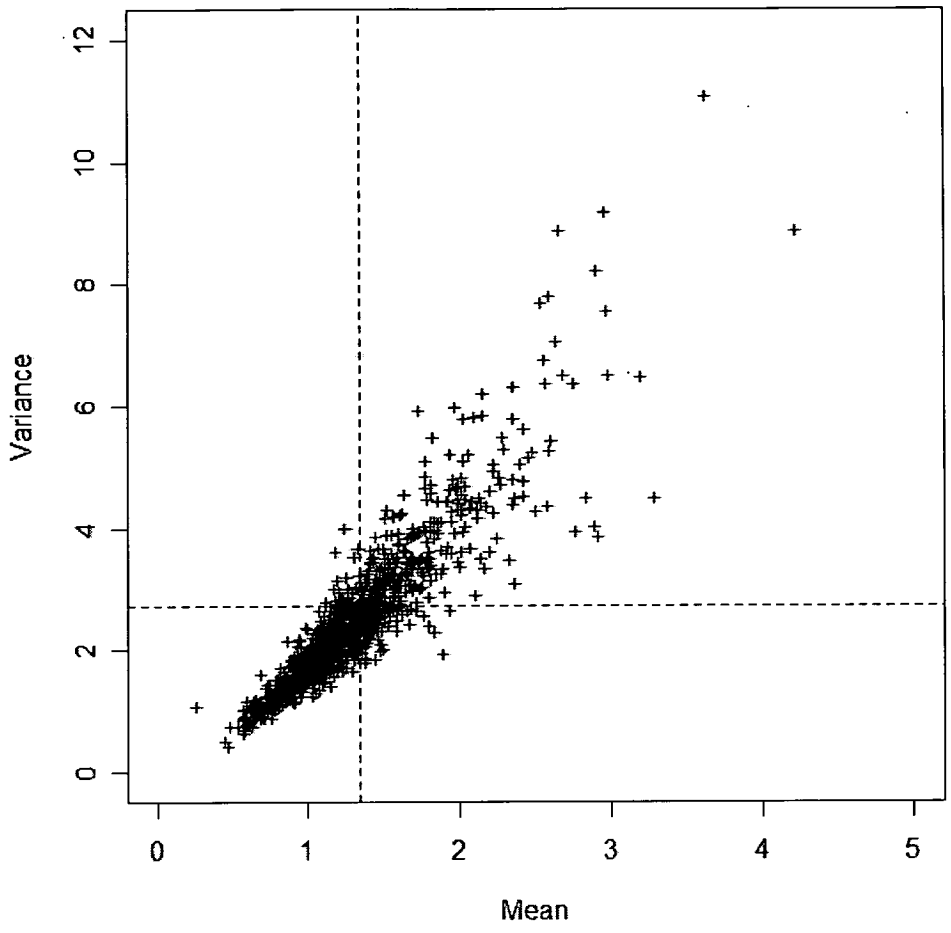

Figure 4.5: Comparison of the mean and variance of the distributions of the maximum test statistic across the chromosome obtained by partial permutation within simulation replicates of Pedigree A with no QTL. The dashed lines indicate the mean and variance of the distribution of the maximum test statistic obtained through simulation.

4.4 Discussion

With the continued search for loci that cause variation in traits of interest, and the required increase in computing power, the classical structured pedigree approaches are being extended to more general cases with the aim of improving the power for both detection and localization of these loci. These approaches also allow the investigation of quantitative genetics in populations that are prohibitive to the use of structured pedigree methods. While the underlying methodology for QTL mapping in general pedigrees is now well developed

(Almasy and Blangero, 1998; George *et al.*, 2000), there is no established methodology for the evaluation of the significance of the results. While this can be directly achieved with the use of simulation, the estimation of the required identity by descent matrix is computationally very expensive and is not practical for large pedigrees.

In this chapter, the extension of permutation methodology to pedigrees with increased complexity was examined. Under the null hypothesis of no QTL, the thresholds estimated by partial permutation showed a downwards bias of 8% and 13% for the two pedigrees examined. When a QTL was present in the simulated data, the direction of the bias was reversed. With simulation Pedigree A, for which partial permutation is measured as being 80% efficient in randomizing the QTL effect, the size of the bias is reasonably small even when a QTL of very large effect was simulated. A QTL explaining 25% of trait variance resulted in an upward bias of 13% in the estimated significance threshold. With simulation Pedigree B, where permutation efficiency is reduced to 63%, the size of the bias in the presence of a QTL of large effect increased to approximately 90%. The size of these biases suggests the partial permutation methodology should be used only in cases where the measured efficiency is high. This excludes all but the simplest extensions of the classically used structured pedigrees, which in turn removes the reasonings behind using more general pedigrees in the first place. It follows that this method will not be useful in estimating significance thresholds for the complex commercial sheep pedigree examined in Chapter 2.

The source of the bias in significance thresholds estimated using partial permutation has been demonstrated to occur through the non-permutable portion of the pedigree. One possible solution to this problem is to split the pedigree into portions that are permutable and non-permutable. As

discussed earlier, while an individual pair may not be directly permuatable, some of its information about inheritance at the locus of interest may be randomized through the permuting of other pairs. Thus, the splitting of the pedigree into permutable and non-permutable parts should not be done at the level of pairs of individuals but at the level of contribution to the test statistic.

While the partial permutation approach to the extension of permutation methodology to general pedigrees did not produce the desired results, other less restrictive approaches may be possible. The restriction on permuting proposed here resulted in partially permuted data sets that all show the same heritability for the trait, or traits, being examined. More flexibility may be obtainable through the generalization of this requirement such that the average heritability over all replicates equals the observed trait heritability. Ideally, this variation would reflect the sampling variation in heritability estimation. An approach showing similarities to this suggestion has been explored by Iturria *et al.* (1999). Their permutation is achieved by arranging the actual trait data to reflect simulated trait data. However, it is unclear whether the thresholds obtained using this method are correct when the parametric model used in simulating trait data does not adequately reflect the actual trait data. Also, the parametric assumptions involved in simulating trait data remove the desirable properties of permutation thresholds. Another approach to permuting trait data could involve the partitioning of actual trait data into permutable and non-permutable components. Such a framework could be based on best linear unbiased predictors (BLUPs) obtained from the variance component framework.

While the extension of the permutation methods would be the ideal solution for the estimation of significance thresholds as this accounts for deviations from distributional assumptions, the generalization of other approaches should also

be considered. Generalization of the estimation of the number of independent tests approach described by Cheverud (2001) is hampered as it is unclear how to calculate the required correlations between markers when these are not directly being transformed for use as an explanatory variable in a linear regression framework. One possibility is to use a measure of the similarity of identity by descent matrices at marker positions such as that used to measure the efficiency of partial permutation. This would require an in-depth study of the statistical properties of such a measure to ensure the 'correlation' it measures was theoretically sound.

The extension of methods that infer the number of independent tests through the examination of the changes of the linkage test statistic across the genome faces several challenges when using a likelihood ratio test for a parameter that is at its boundary under the null hypothesis. The resulting distribution for the likelihood ratio test statistic is a mixture of a point mass at zero and a chi-squared distribution (Self and Liang, 1987). Such a mixture distribution readily decreases the mathematical tractability of the required estimation procedure. Also, the reduction of information given by sequential zero test statistics will result in the reduced accuracy of any estimator that uses the change in the test statistic across the genome.

5 Modelling the Population Dynamics of a Feral Sheep Population: The Soay Sheep of St. Kilda, Scotland

5.1 Overview

The study of feral livestock populations provides an interesting contrast to that of domesticated populations, potentially leading to a better understanding of the genetics of traits including parasite resistance, fertility and various morphological traits. The Soay sheep population of St. Kilda, Scotland is a well studied feral sheep population and provides a valuable resource for such comparisons. In Section 5.2 of this chapter, the archipelago of St. Kilda and the Soay sheep are described followed by an overview of the methods used in the long-term monitoring of the Soay sheep population. Section 5.3 provides a review of the literature on various aspects of the population dynamics of the Soay sheep population. This is followed by the construction of a unified statistical framework for the modelling of the population dynamics (Section 5.4). This framework forms the basis of a simulation model used in the prediction of the linkage disequilibrium structure (Chapter 6) and the effective population size (Chapter 7) of the Soay sheep population.

5.2 St. Kilda and the Soay Sheep

St. Kilda: St. Kilda is the collective name for four islands and several stacks situated in the Atlantic Ocean at 57°49′N, 08°34′W, about 180 kilometres from mainland Scotland and 70 kilometres from the Outer Hebrides (Figure 5.1). The largest island, Hirta, comprises 637 hectares out of the total 840 hectares (Campbell, 1974). From the sea, the only easily accessible point to Hirta is from the south-east, where the hills surrounding the Village Bay provide a somewhat sheltered access to a sandy beach. Hirta forms the centre of a cluster of islands

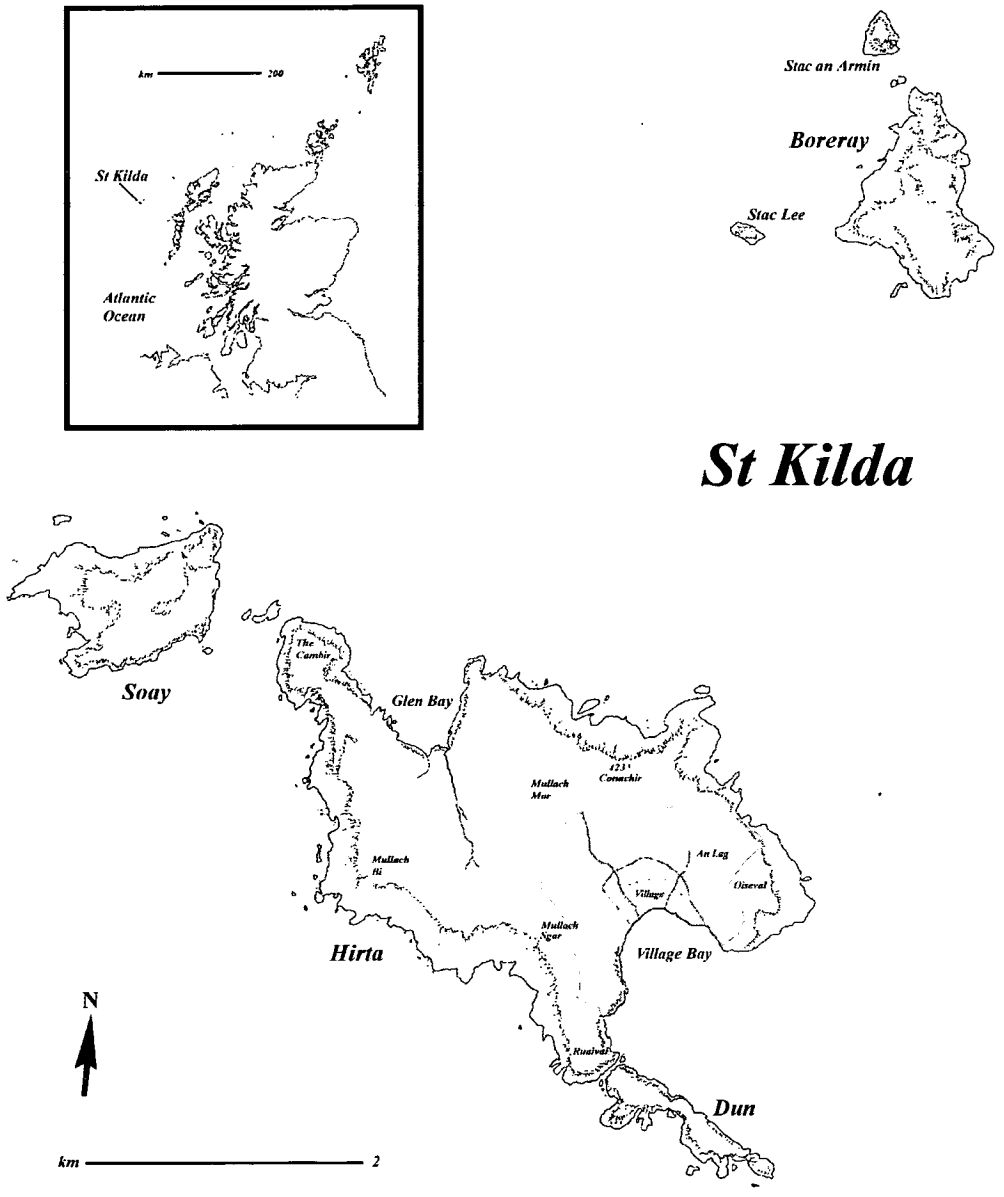

Figure 5.1: Map of the St. Kilda archipelago. The Village Bay study area is shaded. (Derived from Figure 1.1 of Clutton-Brock and Pemberton, 2004.)

with Soay to the north-west (99 hectares) and Dun to the south (32 hectares) while another cluster situated 7km to the north-east is formed by Boreray (77 hectares) and the two major stacs, Stac an Armin and Stac Lee.

The cliff bound islands provide excellent breeding grounds for sea birds, with the vast numbers of fulmars, gannets and puffins nesting there giving St. Kilda one of the largest breeding colonies of sea birds in Europe (Campbell, 1974). The islands also have endemic sub-species of wood mouse (*Apodemus sylvaticus hirtensis*) and winter wren (*Troglodytes troglodytes hirtensis*). St. Kilda became property of the National Trust of Scotland in 1957 (Boyd *et al.*, 1964) and, because of its importance to wildlife, was later designated a National Nature reserve in 1964 and a World Heritage Site in 1986.

The climate of St. Kilda is primarily oceanic although the hills increase the annual rainfall in lowland areas to around 1100-1300mm a year. The islands are buffeted by high winds which occur in every direction throughout the year (Campbell, 1974). The winter climate is highly dependent on the North Atlantic Oscillation (NAO) (Forchhammer *et al.*, 2001). In years where a high pressure system resides over Iceland and low pressure over Azores, winters become wet and windy but are dry and cold when these pressure systems are reversed.

The human occupation of St. Kilda was well documented despite the islands' remoteness. The primary site of human inhabitation was the relatively sheltered Village Bay area where most of the inhabitants' buildings still stand. In 1697 there were 180 inhabitants on the islands but an outbreak of smallpox in 1730 dropped this figure to thirty (Boyd, 1974). By 1800, the population had almost recovered reaching approximately 100 people, but a decrease over the next century began particularly due to a migration of thirty-six people to Australia

in 1856. By 1930 only forty people remained on the island and, due to a lack of physically able people to collect food, the population was evacuated at their own request in August of that year. The majority of the domestic livestock was removed with the people, except the black-faced sheep on Boreray and the Soay sheep on Soay (Boyd, 1974). A substantial effort was made to eradicate all remaining sheep on Hirta, although "at least one blackface survived" (Boyd, 1953).

The Soay Sheep: The Soay sheep are the most primitive domestic breed in Europe, resembling those brought to Britain from mainland Europe about 7000 years ago (Campbell, 1974; Doney *et al.*, 1974). There is no firm archaeological evidence to say when the Soay sheep were introduced to St. Kilda, but it is likely to have been between 1000 and 2000 years ago (Boyd and Boyd, 1990). However, they were eventually restricted to the island of Soay after which they were named.

When compared to improved breeds of sheep, the Soay sheep are small, with narrow bodies, long legs, short tails and narrow faces (Campbell, 1974). Males grow to an average of 53 centimetres in length, while females are slightly smaller growing to 50 centimetres on average (Boyd and Boyd, 1990). The Soay have relatively long leg lengths for their body size in comparison to the Scottish Blackface sheep (Doney *et al.*, 1974). The weight of sheep fluctuates during the year with four-year-old sheep weighing 33 and 23 kilograms in the autumn for males and females respectively, but dropping to 21 and 14 kilograms in the spring (Boyd and Boyd, 1990). Weight gain continues until the sheep are at least five years of age (Doney *et al.*, 1974).

The Soay sheep are close woolled with colours varying from pale buff to black (Boyd *et al.*, 1964). Within these colours it is possible to see a dark and light

morph segregating (Campbell, 1974). A second variant of colour is observed with the majority of sheep having pale-coloured rumps and stomachs (wild-type) while others have a uniform colour (self). From molecularly inferred pedigrees, these variants appear to be segregating in a simple Mendelian fashion at two separate loci, with dark being dominant to light and wild-type dominant to self (Coltman and Pemberton, 2004). The Soay sheep also show polymorphisms for horn development (Doney *et al.*, 1974). The majority of males grow spiral horns with the remaining having small, deformed horns (scurs). Horned females have smaller horns than males. Females also show an additional polled horn variation, with the three horn types being similar in frequency. The current understanding of horn inheritance in the Soay sheep indicates that a single locus with three alleles that are differentially expressed in each sex controls horn type (Coltman and Pemberton, 2004). This is consistent with models suggested for horn inheritance in domestic sheep (Dolling, 1961; Montgomery *et al.*, 1996).

Population Monitoring: In 1932 a flock of 107 Soay sheep was moved from Soay to Hirta for vegetation management at the request of the Marquis of Bute who then owned the islands. The flock consisted of 20 rams, 44 ewes and 43 lambs, of which 21 were female and the remaining castrated males (Boyd, 1953). The population was then allowed to increase while being largely unmanaged. This subsequently provided an opportunity to study a large herbivore in a comparatively simple ecosystem with no migration, predators or competition for grazing from any other species.

There have been two main periods of study on the Soay sheep. In 1955, research started on the ecology of the Soay sheep, with the research intensity being increased between the years 1960 and 1968. The results from this period led to the publication of a book on the Soay sheep of Hirta (Jewell *et al.*, 1974).

In 1985, research on the population dynamics of the Soay sheep was restarted. This coincided with the time when molecular techniques became viable for the accurate monitoring of population ecology. The research activities are concentrated on the Village Bay area (see Figure 5.1) where the catching and sampling of the Soay sheep on a large scale is feasible. The current understanding of the Soay sheep population has been published in a second book (Clutton-Brock and Pemberton, 2004). The methodology described below provides an overview of the current monitoring scheme used on St. Kilda.

Each year three main study expeditions to the islands take place. In late February or early March, a small group visit Hirta for approximately three months. During this time frequent censuses of the study area are performed, recording the distribution of individual sheep throughout the area. Additional searches are performed for dead sheep, which have various anatomical samples removed and death date recorded. Approximately half-way through this expedition the lambs are born. A substantial effort is made to catch all lambs, which then have their ears tagged and sex, coat colour and mother's identity recorded. The small hole punch used in the tagging of the lambs provides a biological sample for genetic analysis.

The second expedition occurs in July and August. Initially, a small team visits Hirta to carry out more censuses of the study area. Later, a larger team attempts to fence off the entire study area and catch all sheep within it. All animals caught are recorded and have a variety of measurements taken including weight, leg and horn measurements and levels of parasitic infection. In addition, a census of the whole of Hirta is performed by three teams using designated census routes that have been followed since 1959 (Boyd *et al.*, 1964). In the census the sheep are classified by their age class (lamb or adult), sex and coat type where

possible.

The final study expedition occurs between October and December with the objective of collecting information on rutting behaviour. As with other study periods, censuses of the study area are performed. Information is also collected on when individual ewes come into oestrus. Any untagged rams entering the study area are immobilized, tagged, measured and have blood samples taken. During this time the movements of selected individual males are followed and any subsequent formation of consorts of tagged ewes are noted.

5.3 Population Dynamics of the Soay Sheep

Population Size and Composition: The population size of the Soay sheep fluctuates widely between years, with island census totals varying from 610 to 1968 (Figure 5.2). The local maxima occur at irregular intervals indicating an aspect of randomness to these fluctuations. Given the similar fluctuations of sheep populations on surrounding islands, the major diving factor of these fluctuations appears to be from interrelated climatic and nutrition variables (Grenfell *et al.*, 1998). This is supported by the observation that malnutrition, caused by overpopulation and lack of grass growth, is the major cause of death in years of high mortality. Boyd (1974) showed there was a weak upward trend in population size in the years until 1973 and this increase appears to have continued through to the present. Although the reasons for this increase are unknown, possible explanations are the adaptation of the sheep to their surroundings or, more likely, the improvement of counting methods. For the counts taken in 1965 and 1966 approximately 73% of rams, 83% of lambs and 90% of ewes were counted (Grubb, 1974). There may also be an effect due to the changing of island census time from May to August in recent years.

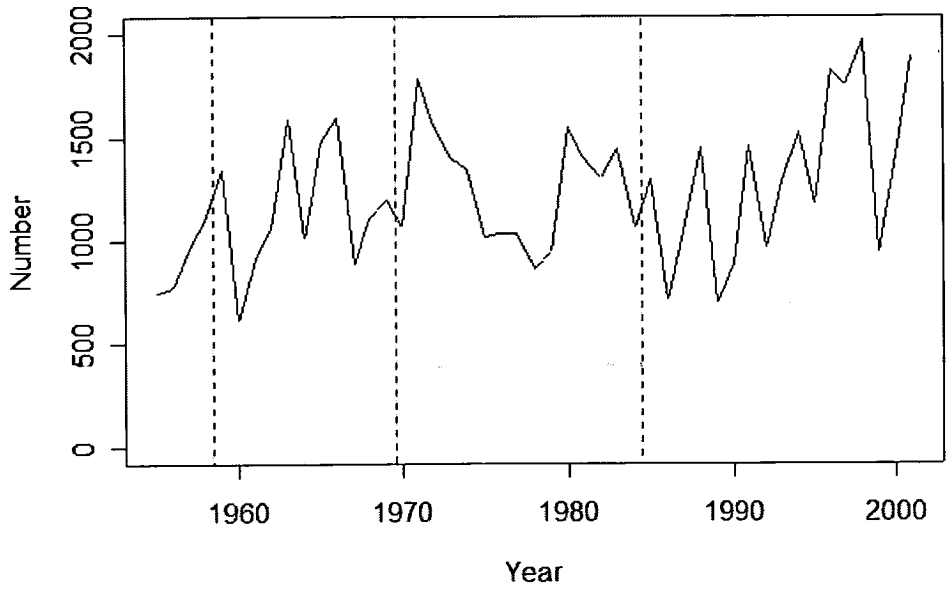

Figure 5.2: Plot of census population size of Soay sheep on Hirta. Counts were performed by research groups in the periods 1959 to 1969 and 1985 to 2001. Other years (shaded) were carried out by the reserve wardens.

Figure 5.3 compares aspects of the Village Bay population with that of the entire island. The census counts in the Village Bay area show a strong correlation with the island census counts (Figure 5.3a and 5.3b, $r = 0.962, p < 0.001$), highlighting that a fluctuation in population size is not area specific but affects the whole island. The proportions of lambs in the two populations are similar (Figure 5.3c), although a slightly higher proportion of lambs is given by the island census. The adult sex ratio is higher in the Village Bay population (Figure 5.3d), but this is likely to be due to the increased accuracy of the Village Bay count compared with the island census, as rams are under-counted as opposed to the ewes in the island census (see above).

Both proportion of lambs in the population and the adult sex ratio vary over time (Boyd, 1974). The proportion of lambs in the population shows only moderate synchrony with the population size (Figure 5.4). In contrast, the adult sex ratio fluctuates in synchrony with the population size with relatively less

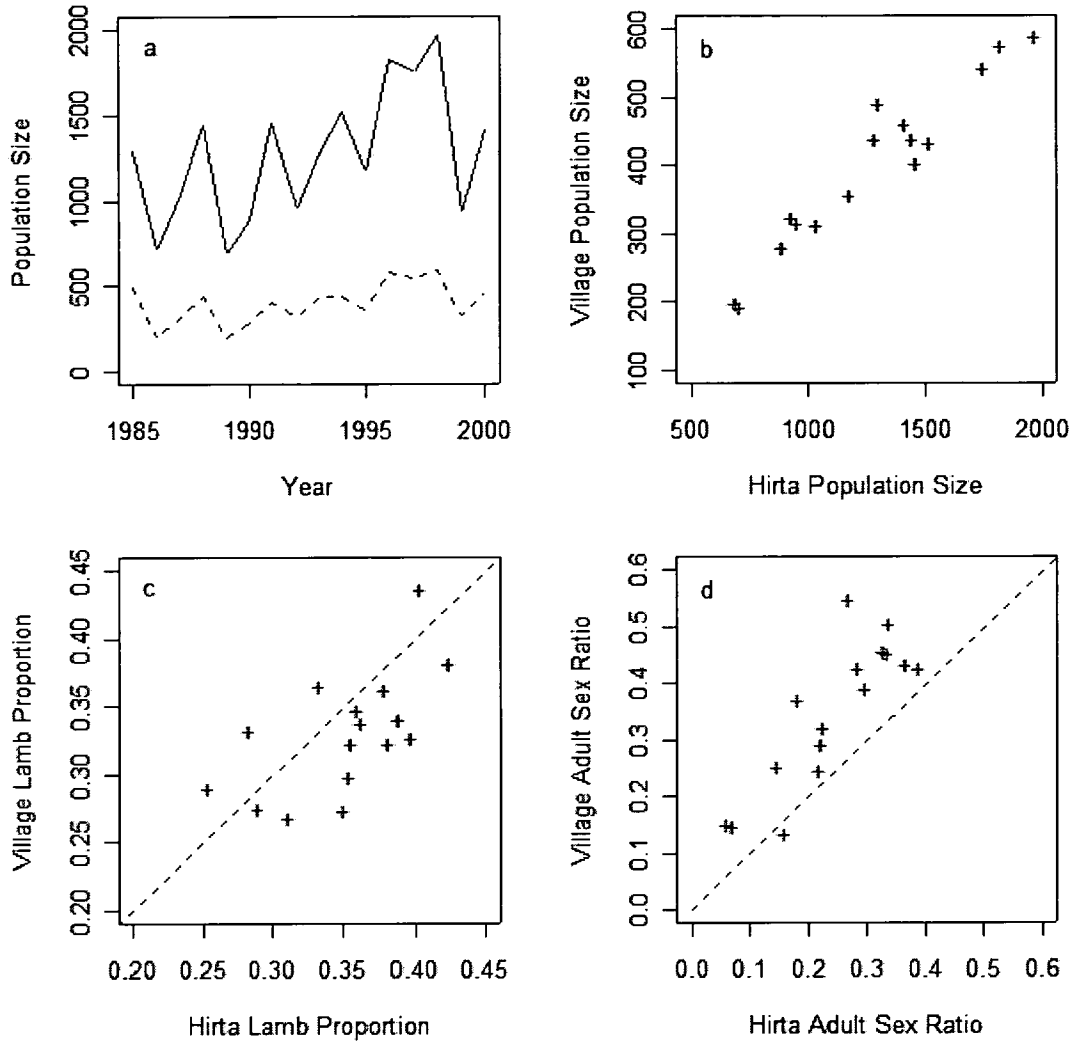

Figure 5.3: Comparisons between Village Bay and Hirta population data. (a) Plot of census population size of Hirta (solid line) and the Village Bay population (dashed line). (b) Plot comparing Hirta and Village Bay population sizes. (c) Comparison of the proportion of lambs in the two populations. The dashed line indicates a perfect relationship (d) Comparison of the adult sex ratio (males/females) in the two populations.

rams being present in years of low population size (Figure 5.5). It follows that the male population is more susceptible to starvation and thus the fluctuations in male numbers are of greater magnitude relative to that of females.

Survival: The survival of sheep to the following year depends on a variety of factors including sex, age, weight and population size. In the adult population, almost all deaths occurs between January and April (Clutton-Brock *et al.*, 1997). The primary cause of death appears to be starvation, as observed through the depletion of fat reserves, although this may be mediated by increased susceptibility to parasitic infection (Gulland, 1992). It follows that an increased fat reserve, and thus higher weight, reduces the risk of death. As juvenile sheep have lower fat reserves, they tend to die earlier than adult sheep. The risk of death for juvenile sheep is partially mediated through their mother's condition, with very young and old ewes as well as lighter ewes tending to give birth to lighter lambs. Juvenile females are more likely to survive the winter than juvenile males, as are sheep born later in the year (Clutton-Brock *et al.*, 1992). The effect of these predicting factors for sheep deaths is increased in years of high density (Milner *et al.*, 1999). The reproductive status of female sheep is a further explanatory factor, with pregnant sheep at higher risk of dying. This effect is particularly noticeable in juvenile females. However, any effect of pregnancy does not carry over to subsequent years (Clutton-Brock *et al.*, 1996).

Figure 5.6 shows average survival curves for male and female sheep with known fates. Although the survival up to two years old appears independent of sex of the sheep, thus conflicting with the discussion above, care needs to be taken in the interpretation. Important aspects of population dynamics, such as the observed increase in the proportion of males born in years of low population density (see below), are averaged across years obscuring some variation in

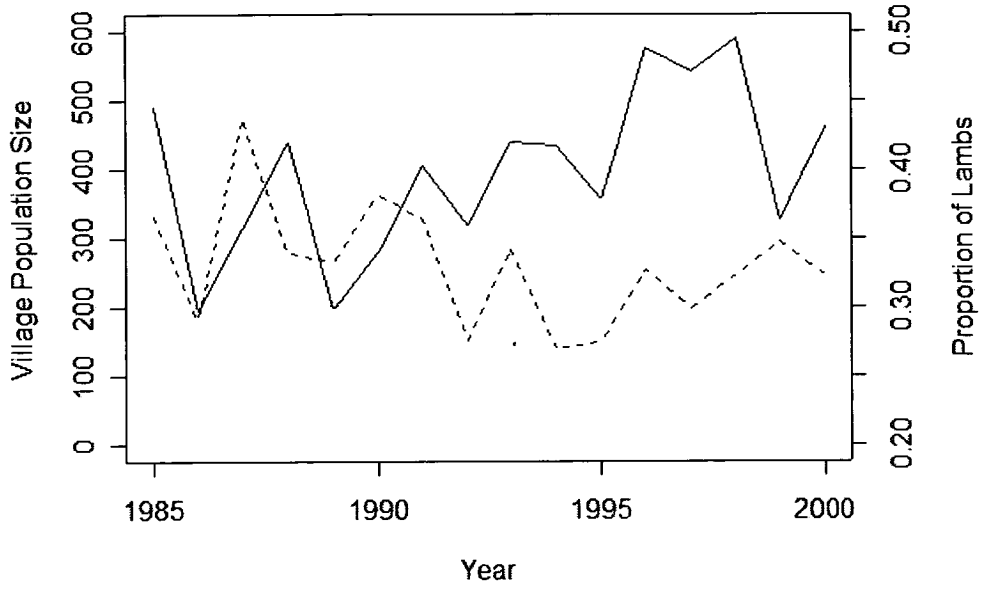

Figure 5.4: Plot comparing Village Bay population size (solid line) and the proportion of lambs (dashed line) in the the Village Bay population.

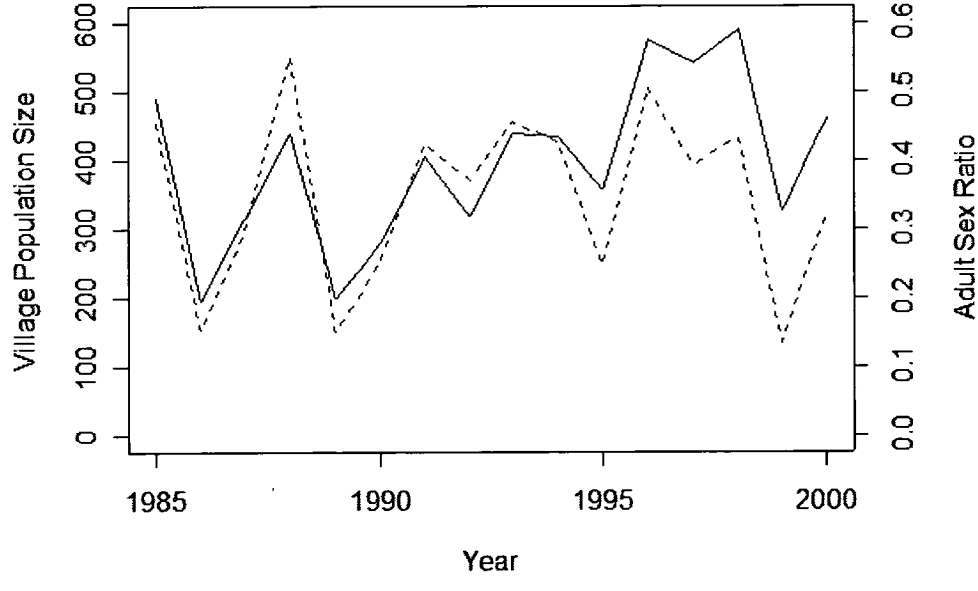

Figure 5.5: Plot comparing Village Bay population size (solid line) and the male to female sex ratio (dashed line) in the adult sheep of the Village Bay population.

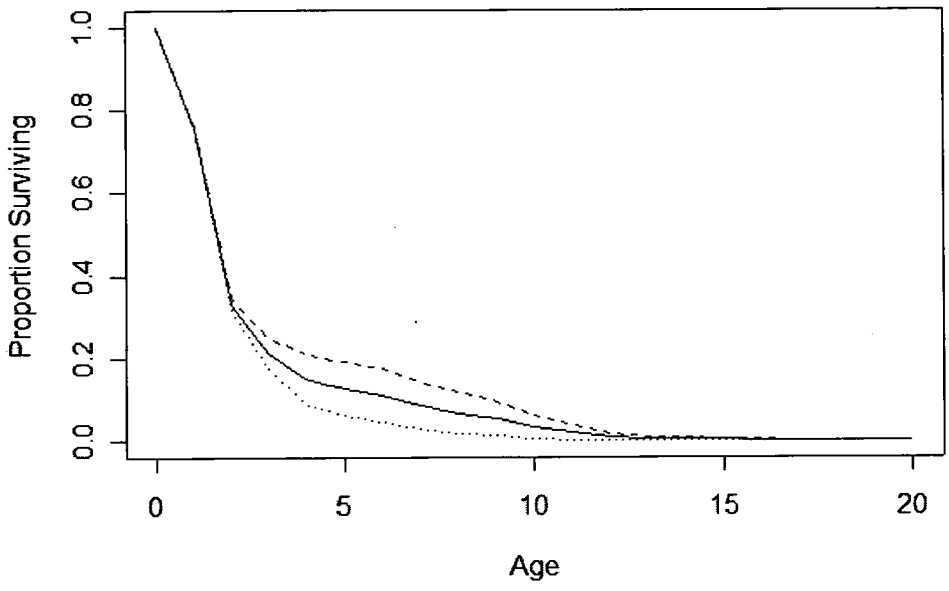

Figure 5.6: Plot showing the empirical survival probabilities for sheep with known fate. The survival curves represent the overall survival (solid line), survival in males (dotted) and survival in females (dashed).

survival. However, some general aspects of survival in the Soay sheep become apparent. Over half of the sheep fail to live beyond two years of age and it is at this point that differences in cohort sex ratio become fully apparent. No males were observed to live longer than twelve years and no females longer than sixteen.

Reproduction: In November, almost all Soay females come into oestrus for a period of one to four days and continue to cycle at fifteen day intervals until fertilization has been achieved (Clutton-Brock *et al.*, 2004b). It is rare for adult females not to produce lambs, although fecundity decreases in ewes over ten years of age. It is common for conception to occur in the first year of life when the juvenile sheep are about seven month of age. This is dependant on population size with up to 81% of juveniles conceiving in year of low population density and this reduces to 6% with high population density (Clutton-Brock *et al.*, 2004a). Twinning rates in the Soay sheep vary with population size and weight of the ewe. In years of high population density, the proportion of twins

is low compared to years of low density, the overall rate varying from 2% to 20% (Clutton-Brock *et al.*, 1991). Heavier ewes are more likely to produce twins than lighter sheep. Also, juveniles and older ewes show lower twinning rates than middle aged ewes. The proportion of male offspring in a year ranges from 43% to 56% and is weakly associated with the previous years population density (Lindström *et al.*, 2002).

There are two primary male reproductive strategies, with rams either forming a consort pair with a ewe and then guarding the ewe for as long as possible or, in the case of less dominant rams, attempting to get matings by running at the ewe, referred to as coursing, which often results in the ewe being chased and mated by multiple different males (Stevenson *et al.*, 2004). The mating strategy used is dependant on the ewes body weight, with rams preferentially guarding ewes of greater weight. The promiscuity of Soay ewes is evident in the analysis of paternities of twins in which only 21 out of 80 (26%) twin pairs examined had the same father (Pemberton *et al.*, 1999). The average male breeding success decreases with increasing population density, primarily due to the smaller proportion of females in years of high population density. This effect is more pronounced in adult males (Pemberton *et al.*, 2004), although this is due to juveniles and yearlings producing a relatively small proportion of lambs. However, adult males are always more likely to produce at least one offspring. Male reproductive success is highly skewed, with paternities assigned at greater than 80% confidence estimating distribution of male lifetime breeding success as having a mean of 1.05 and standard deviation 2.10 (Pemberton *et al.*, 2004). The population density in which a male lamb is born provides a major explanatory factor for his lifetime breeding success, with lambs born into high population density having low survival and unsuccessful first ruts compared to lambs born into low population density.

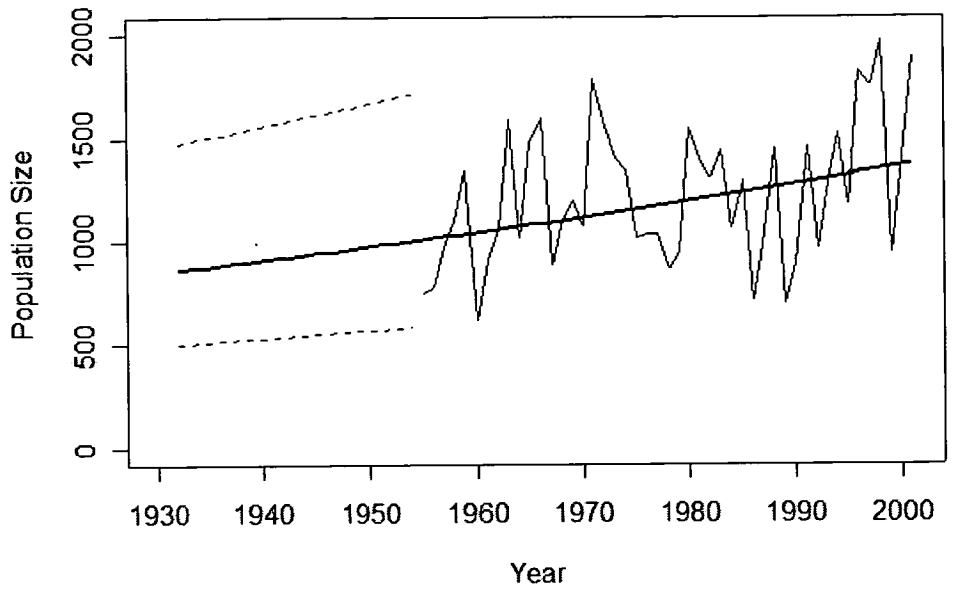

Figure 5.7: Plot of census population size showing trend (bold) and approximate 95% confidence intervals for predicted population size in years prior to census taking (dashed).

5.4 Statistical Modelling of Population Dynamics

Population Size: The model for the fluctuations in population size is based on the log of the census population size as this allows naturally for the indication of density dependant selection in the population (Grenfell *et al.*, 1992). This also removes the potential confounding of the fluctuation size with the proportion of sheep being counted. The trend for increase in population size was removed by considering the residuals of the model

$$\ln n_i = -6.362 + 0.00679 y_i + \epsilon_i \tag{5.1}$$

where n_i is the census population size of year y_i and ϵ_i is normally distributed with standard deviation 0.269 (see Figure 5.7). The residuals of this model standardized to have a variance of one for use in all following models and are referred to as residual population size.

The changing residual population size was modelled using a self-exciting threshold autoregressive (SETAR) model (Tong, 1990), with model selection based on the improved cross-validation criterion, C_u, proposed by de Gooijer (2001). SETAR models are useful in modelling data in which future values depend on previous values but this dependence varies with the current value of the system. Each region where a different model is applied in predicting future values is called a regime, with its boundaries known as thresholds. In each regime an auto-regressive (AR) model is used in predicting future values. This type of model has previously been applied to the Soay sheep population (Grenfell *et al.*, 1998) but without removing the trend and using a different model selection criterion.

The model selection procedure found no evidence for the existence of more than one regime in the SETAR model thus reducing model selection to that of a standard autoregressive model. Furthermore, no auto-correlation or partial auto-correlation is observed between successive residual population sizes and these show no deviation from normality. This indicates a simple model of normally distributed noise is appropriate for residual population size.

Model Selection Procedure: A modified model selection procedure was used for modelling the proportion of lambs, adult sex ratio and survival. All models are based on linear regressions on residual population sizes of the current and previous years. Initial models used residual population sizes from the previous five years, with this being successively reduced when the earliest year in the model was not significant at the 5% level. This modified selection procedure allows effects for years after the earliest significant year to be non-significant. As a delayed response to population size is unlikely, this type of model selection is useful, especially when the power to detect small effects is low. One disadvantage

of this selection regime is the potential of false positive results that lead to a highly over-parameterised model, but choosing reasonable starting models reduces the likelihood of this occurring.

Lamb Proportion: The proportion of lambs, p_i, in the population was modelled using a logistic regression model of the form

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \mu + \sum_{j=i-d}^{i} \beta_j r_j + \epsilon_i \tag{5.2}$$

where d defines the delay parameter that is successively reduced in the model fitting procedure and ϵ_i is normally distributed noise. The model selection procedure found no evidence for a relationship between residual population size and the proportion of lambs in the population. As the untransformed proportions showed no significant deviation from normality a simple model of the proportion of lambs being normally distributed is most parsimonious. The mean and standard deviation estimated from the data are 0.331 and 0.0442 respectively.

Adult Sex Ratio: The ratio of the number of adult males to adult females, s_i, was modelled as

$$\ln(s_i) = \mu + \sum_{j=i-d}^{i} \beta_j r_j + \epsilon_i \tag{5.3}$$

where the parameters as defined above. The model selection procedure obtained d equal to three, reduced from an initial value of five, giving

$$\ln(s_i) = -1.145 + 0.385 r_i - 0.121 r_{i-1} - 0.0767 r_{i-2} - 0.110 r_{i-3} + \epsilon_i \tag{5.4}$$

where the error term is normally distributed with standard deviation 0.239.

Survival: For the modelling of survival, only sheep with known birth date and

known fate (either a known death date or known to be still alive) were used. Separate models for the effect of residual population size on survival were made for each age and sex. For each year, the hazard function was evaluated using the Nelson-Aalen estimator

$$h_{i,j,s} = \frac{d_{i,j,s}}{n_{i,j,s}} \tag{5.5}$$

where $d_{i,j,s}$ is the number of observed deaths in the $n_{i,j,s}$ sheep of age i and sex s alive in year j and followed to year $j+1$ (Nelson, 1972; Aalen, 1978). The effect of residual population size on the hazard function was modelled using a weighted regression of the form

$$\text{logit}(h_{i,j,s}) = \ln\left(\frac{h_{i,j,s}}{1-h_{i,j,s}}\right) = \mu + \sum_{j=i-d}^{i} \beta_j r_j + \epsilon_i \tag{5.6}$$

where d was selected using the model selection procedure described above starting from an initial value of four and weights were given by $n_{i,j,s}$. Coefficients for the linear models are given in Table 5.1. In older age groups, where there was data available from only a few years and there was no observable effect of residual population size on the hazard function, the hazard was modelled as a Beta distribution with parameters $d_{i,s}+1$ and $n_{i,s}-d_{i,s}+1$, where $d_{i,s}$ and $n_{i,s}$ are obtained from the summation over j of $d_{i,j,s}$ and $n_{i,j,s}$ respectively (Table 5.2). As no observations of males older than twelve years or females older than sixteen years have been made, these were considered the upper limit for the lifespan of the Soay sheep.

5.5 Discussion

The long term monitoring of the Soay sheep population provides an opportunity to understand the dynamics of population size and how this effects population composition and individual life histories. In this chapter, a unified framework of statistical models was constructed in order to provided the basis for a simulation

Table 5.1: Parameters for linear models for survival in the Soay sheep (given to three decimal places).

Sex	**Age**	μ	β_j	β_{j-1}	β_{j-2}	σ
Male	0	-1.477	-	-	-	1.772
	1	-0.362	-1.291	1.535	-	2.556
	2	-1.494	-1.332	0.818	-	2.705
	3	-1.510	-1.448	0.484	-0.948	1.529
	4	-1.217	-0.697	-	-	1.396
	5	-1.133	-0.775	-	-	1.530
	6	-0.860	-0.864	-	-	1.059
Female	0	-1.776	-	-	-	4.270
	1	-0.694	-0.792	1.086	-	5.014
	2	-2.333	-1.017	0.364	-	3.265
	3	-2.910	-0.784	0.378	-0.758	2.382
	4	-2.390	-0.492	0.352	-	1.797
	5	-2.641	-0.539	-	-	3.784
	6	-2.317	-0.512	-	-	2.930
	7	-2.022	-0.687	-	-	2.118
	8	-1.901	-0.958	-	-	2.027
	9	-0.883	-1.016	-	-	2.059
	10	-1.170	-0.948	-	-	3.261

Table 5.2: Values used to calculate parameters for Beta distribution survival models in the older sheep age classes.

Sex	Age	d_i	n_i
Male	7	32	77
	8	14	41
	9	6	21
	10	4	11
	11	3	6
Female	11	20	81
	12	12	44
	13	11	30
	14	6	14
	15	1	4

model of the Soay sheep population.

The modelling of the large fluctuations in population size found no evidence for the dependency of the current population size on those of previous years. Grenfell *et al.* (1998) modelled the population size similarly but did not remove the trend and based model selection on Akaike information criterion (AIC). Under these conditions, the data was separated into two regimes. Below the threshold, the population size for the following year was given by a simple recurrence relationship with the current population size showing an increase in numbers with normally distributed noise. Above the threshold, there was no significant relationship between the current population size and that of the following year. As no attempt was made to correct for the observed increase in overall population number over time this will be confounded with the linear effect identified in the lower regime. Another difference was that

model selection was based on AIC, which tends to over-parameterize the final model (de Gooijer, 2001) and should be used cautiously as the exact likelihood of a time series generated by a SETAR process does not exist (Tong, 1990). The conclusion of the study by Grenfell *et al.* that correlations between sheep population sizes on the islands of St. Kilda exist due to the influence of weather conditions on survival would be strengthened if the simpler model of population size is used. However, it is unlikely that no dependence exists in the observed values of population size as vegetation growth, and thus the susceptibility of the population to starvation, is regulated through weather conditions that show cyclic patterns on the scale of several years. Also, the weather during winter directly effects survival probabilities (Coulson *et al.*, 2001).

The model selection procedure for the adult sex ratio detected effects of population size up to three years previous. The majority of the effect of population size on adult sex ratio is through the current population size, as seen by the coefficients for the three previous years being approximately half of the error standard deviation. Thus residual population size of the previous three years would need to have remained relatively constant in the extremes of the distribution for residual population size in order to produce a readily observable effect. Interpretation of the model based solely on the coefficient of current population size shows the ratio of male to female sheep increases with increasing population size. This is consistent with the observation that males are more susceptible to starvation in winters where mortality is high.

No effect of population size was observed on the proportion of lambs in the population. This result initially appears to contradict the observation of a higher proportion of adult females in years of low population size that would possibility indicate a larger proportion of lambs. However, this effect is reduced through

the increased mortality of pregnant females compared to those that have not conceived (Clutton-Brock *et al.*, 1996) and the high neonatal mortality that follows high overwinter mortality (Clutton-Brock *et al.*, 2004a), although it is unlikely these factors cancel out completely. It is probable that a relationship between residual population size and the proportion of lambs does exist and the detection of no significant effects is due to the low power afforded by the relatively small data set.

The models for survival give some insight to the life-history of the Soay sheep population. The results indicate that in their first year of life survival through the winter does not depend on population size. This is contradictory to the observation that the effects of various factors that influence survival, such as weight, mother age and birth date, are increased in years of high mortality (Milner *et al.*, 1999). This may be a consequence of using a linear model framework in modelling effects that are likely to deviate from linearity. The coefficients of the effect of residual population size all change the estimated hazard in the same direction for all age groups. The negative coefficients of the effect of the current residual population size on survival show that, as expected, the hazard is increased in years with a high residual population size and decreased when residual population size is low. The observation of parameter estimates all being of the same sign and similar magnitude indicates the survival models are over parameterized. However, from a biological perspective it is not obvious how to group different ages as different ages would be expected to show some differences in survival. Thus, all ages should have survival modelled separately. The survival of rams older than six years of age and ewes older than ten showed no effect of population size. This is primarily due to the relatively small numbers of sheep that survive to these ages giving a lack of power to detect effects of residual population size. Also, obtaining such a result entirely due to

the effects of ageing is unlikely, as these would be expected to be amplified in years of high population size.

The simple models presented here provide a strong basis for the investigation of the genetical properties of the Soay sheep population through the use of simulation. In the following chapters, such a simulation model is developed and used to investigate the linkage disequilibrium structure of the population (Chapter 6) and the effective population size of the Soay sheep population since its introduction to Hirta (Chapter 7).

6 Modelling of the Linkage Disequilibrium Structure of the Soay Sheep Population

6.1 Introduction

The long-term monitoring of the Soay sheep population provides an opportunity for the understanding the quantitative genetics of various traits associated with fitness, including lifetime breeding success, parasite resistance and survival. Also, traits such as birth weight can be compared to domestic populations, providing an insight into the effects of population management on the expression of quantitative traits. The partial pedigree reconstruction that has been achieved while developing an understanding of the dynamics of male reproductive success (Pemberton *et al.*, 1999) can be used in a variance component linkage analysis (Amos, 1994; Almasy and Blangero, 1998; George *et al.*, 2000) to detect quantitative trait loci (QTL) segregating within the population. However, although this approach is useful for the initial detection of regions of the genome that are of interest, the incomplete nature of the pedigree leads to information from many individuals in the population being unused.

A second approach to locating genetic variants underlying quantitative traits is linkage disequilibrium or association mapping. Linkage disequilibrium (LD) is the non-random association of alleles at two (or more) loci. Both human and livestock geneticists have suggested that this association can be used to fine map QTL (e.g. Goddard, 1991; Terwilliger and Weiss, 1998; Haley, 1999) by testing for an association between marker alleles and the trait of interest in a large sample of individuals. However, before implementing an LD mapping scheme it is important to know the extent of LD in the population so the appropriate marker density can be used. The overall pattern of the decay of LD with genetic distance and variation about this decay is referred to as the structure of LD of a

population.

Recently, investigations have been undertaken into the LD structures of several domestic livestock species, including cattle (Farnir *et al.*, 2000; Tenesa *et al.*, 2003), pigs (Nsengimana *et al.*, 2004) and sheep (McRae *et al.*, 2002). These studies showed a significant decay of LD with distance, a requirement for LD mapping. However, there is a large amount of variation about the average decay with some tightly linked marker pairs showing no significant LD. Also, significant amounts of LD were observed between unlinked marker pairs. This impedes using LD mapping in these populations as (false) positive results may occur in regions unlinked to any QTL and negative results in tightly linked regions. Despite these drawbacks, the potential of LD mapping has been realized in at least two cases in dairy cattle, with the fine mapping of a twinning locus to a region of less than 1cM (Meuwissen *et al.*, 2002) and the identification and subsequent characterization of a gene for milk yield and composition (Farnir *et al.*, 2002; Grisart *et al.*, 2002).

Despite the similarity between the LD structures of domestic livestock populations, it is necessary to investigate the amount of LD in the Soay sheep population due to its different population history. This has been demonstrated by the variety of LD structures in human populations (reviewed in Boehnke, 2000; Pritchard and Przeworski, 2001; Ardlie *et al.*, 2002; Wall and Pritchard, 2003). For example, the length that useful levels of LD extends in the human genome has been observed to be as small as 3-5kb (Dunning *et al.*, 2000) or over 1Mb (Taillon-Miller *et al.*, 2000). Also, although some studies found increased levels of LD in admixed populations (Wilson and Goldstein, 2000; Reich *et al.*, 2001), others have shown similar levels of LD in isolated and mixed populations (Eaves *et al.*, 2000; Taillon-Miller *et al.*, 2000).

Direct evaluation of the LD structure of a population is both time-consuming and expensive. Theoretical estimation of LD levels in a population is complex due to the large number of inter-related factors involved in the formation of LD, which include genetic drift, population admixture and natural selection. Simulation studies provide a useful alternative for the evaluation of LD structure. This approach has been used in humans (Kruglyak, 1999) and these predictions for the level of LD are consistent with those observed from experimental data in some populations (Dunning *et al.*, 2000), despite the very simple model applied. However, the wide range of LD structures observed in human populations serves as a warning when generalizing simulation results to populations other than that intended.

In this chapter, the LD structure of the Soay sheep population is investigated through the use of a simulation model. The understanding of the population dynamics of the Soay sheep presented in Chapter 5 provides a strong basis for such a simulation and allows the construction of a model that should predict the overall structure of LD in the population with reasonable accuracy.

6.2 Methods

Simulation Model: The Soay sheep population was modelled mimicking their history since the introduction to Hirta in 1932. For each simulation replicate, the initial haplotypes, which contain a pair of loci separated by a random distance between 0 and 40cM, were simulated using a coalescent process (see below). The ages of the adults in the original sheep were sampled from the equilibrium age structure given by the average survival estimates for each sex. The castrated ram lambs that were initially transferred to Hirta were not included in the simulation, as these would not leave any descendants. The initial population

growth on Hirta was modelled by the doubling of the population size each year until this exceeded the population size trend given in Equation 5.1. From this point, until island census data was available, the population size was given by a realization from the model for population size developed in Chapter 5. For each year that the simulation iterated, realizations from models for sex ratio, proportion of lambs and survival were taken, conditioned on the realised residual population size. Given the observed yearly pattern of mortality of the Soay sheep after the rut but before the birth of lambs and the fact that rams can have offspring born after their death but ewes cannot, the following steps were cycled through in each year of the simulation: (1) Death of ewes; (2) Creation of lambs from remaining sheep; (3) Death of rams; (4) Aging of sheep which remain from the previous year. The number of sheep that were to die at each step of the simulation model was determined by the required adult sex ratio and proportion of lambs given the residual population size. The survival models were used only to give an appropriate age structure to the population. This was achieved by choosing a random sheep and comparing a random number to the realized hazard for sheep of that age, h_i. If the random number was less than h_i, the sheep was removed from the population, otherwise it was returned to the sampling pool. This was repeated until the required population size was achieved. The efficiency of the steps involved was greatly increased through the use of rejection sampling methodology where h_i is replaced by $h_i / \max(h)$.

The approximate uniformity of female reproductive success was modelled by assigning lambs to individual ewes or, if the number of lambs was greater than the number of ewes, all ewes were assigned one lamb with the remaining lambs being distributed to individual ewes. The observation that the proportion of births producing twins decreases with population size is not modelled directly as this association is intrinsically modelled through the changing population size.

Male breeding success was modelled by assigning each male individual a random success variable from a two-parameter gamma distribution with mean of 1.05 and standard deviation of 2.10 as estimated from inferred paternities (Pemberton *et al.*, 2004). As with survival, the selection of male parents was achieved using rejection sampling methodology. For ewes having a single lamb, a male parent was chosen by randomly selecting a male and comparing their success variable (scaled to have maximum value one) to a uniform random number. The male was selected when the random number was less than the scaled success variable; otherwise the sampling process was repeated. When a ewe was assigned as having twins, firstly the number of male parents was randomly determined, with 26% being assigned to one ram (Pemberton *et al.*, 1999). The sampling of male parents for twins with different fathers was achieved using the same methodology as for single lambs. For sampling a single father for twin lambs, the square of the scaled success variable was compared to the random uniform number to correct for the fact the selected individual was being assigned two paternities. Haplotypes passed on to lambs were generated assuming recombination rates given by Haldane's mapping function.

Generation of Initial Haplotypes: Although the population fluctuations on Hirta and its surrounding islands have been shown to be in synchrony (Grenfell *et al.*, 1998) and estimates about the period of time that the Soay sheep have been present on Soay are available (Boyd and Boyd, 1990), it is difficult to extend these concepts to generating appropriate haplotypes for the initial sheep that were moved to Hirta. However some of the properties of the current Soay sheep population can be used to provide an appropriate bench mark for the simulation of these haplotypes. Specifically, they can be used to estimate the effective population size of the original population on Soay.

To estimate the genome wide heterozygosity of the Soay sheep population, a panel of 144 microsatellite markers with a high polymorphic information content (PIC) was genotyped in five rams chosen for their high prolificacy (J. Pemberton, personal communication). This gave an estimate of the genome-wide heterozygosity of 0.385 (SE = 0.009). Simulating the population starting with founder chromosomes having uniquely identifiable alleles indicates that heterozygosity in the Soay sheep has decreased by 7.1% since their introduction to Hirta (see Chapter 7). Thus the average heterozygosity of the microsatellite markers examined in the founding sheep is inferred to be 0.414. Under a neutral model, the heterozygosity at a locus equals $\theta/(1+\theta)$, $\theta = 4N_e\mu$ where N_e is the effective population size and μ is the mutation rate at the locus (Kimura and Crow, 1964). A reasonable estimate for μ will be in the high end of the observed mutation rate of microsatellite loci in sheep (Crawford and Cuthbertson, 1996) given the highly polymorphic loci chosen to measure the average heterozygosity. Using $\mu = 0.005$ gives $N_e = 35.3$ which is approximately 15% of the average population size on Soay. This is consistent with estimates of N_e given in Chapter 7 and within the range of estimates of N_e which account for fluctuations in population size, variation in family size and unequal sex ratio in other populations (Frankham, 1995a).

For each replicate, 40Mb of sequence was simulated with a neutral coalescent model using the program "ms" (Hudson, 2002). This program assumes an infinite-sites model of mutation, generating loci with histories similar to SNP markers. The rate of mutation was assumed to be $\mu = 10^{-8}$ mutations per base pair per generation, which is consistent with estimates of sequence mutation rate in mammals (Nachman and Crowell, 2000). Recombination was assumed to occur at a rate of $r = 10^{-8}$ between base pairs giving a 1cM/Mb ratio, similar to that observed on average in the sheep genome (Maddox *et al.*, 2001). Along

with the effective population size estimate given above, these values were used to calculate the input parameters $\theta = 4N_e\mu$ and $\rho = 4N_e r$. The program generated a large number of polymorphic sites (mean 320) of which the left most and another random site were chosen for use in the remaining simulation, thus giving an approximately uniform distribution of allele separation across the simulation replicates. The average number of polymorphic sites is consistent with that obtained theoretically (Watterson, 1975).

Measuring Linkage Disequilibrium: There are several measures available for the measurement of the amount of linkage disequilibrium between a pair of loci, each having its own advantages and disadvantages (Hedrick, 1987; Lewontin, 1988; Devlin and Risch, 1995). LD was measured using the standardised measure, D' (Lewontin, 1964), as this allowed for comparison with previous studies of LD in livestock which have used a multiallelic extension (Hedrick, 1987) of this measure. D' is calculated as

$$D'_{ij} = \frac{D_{ij}}{D_{max}} \tag{6.1}$$

where

$$D_{ij} = x_{ij} - p_i q_j \tag{6.2}$$

and

$$D_{max} = \begin{cases} \min[p_i q_j, (1-p_i)(1-q_j)]; & D_{ij} < 0 \\ \min[p_i(1-q_j), (1-p_i)q_j]; & D_{ij} > 0 \end{cases} \tag{6.3}$$

where x_{ij} is the frequency of gametes with both allele i at the first marker and allele j at the second marker and p_i and q_j are the frequencies of allele i at the first marker and allele j at the second marker respectively. As the sign of the D' statistic depends on the arbitrary choice alleles at each locus, its absolute value is used throughout this chapter. The D' statistic has a constant range over all allele frequencies, taking a value of 0 when the loci are in linkage equilibrium

and a value of 1 when the loci are in complete disequilibrium. It should be noted that complete LD is different from perfect LD where alleles have a one-to-one relationship. As D' can be biased upwards when one or both markers contain rare alleles (Eyre-Walker, 2000), replicates giving final minor allele frequencies less than 0.1 were repeated.

As the statistical properties of the D' measure of LD are not well understood, the primary simulation results are also presented using the r^2 measure calculated as

$$r_{ij}^2 = \frac{D_{ij}^2}{p_i(1-p_i)qj(1-q_j)} \tag{6.4}$$

using notation as above. The most important statistical property of this measure is the relationship between r^2 and the chi-square contingency table test (Hill and Robertson, 1968). Using this property, it can be shown that the power to detect a trait locus in LD with marker locus is inversely proportional to the value of r^2 between these loci (Pritchard and Przeworski, 2001). The expected value of r^2 is also related to the effective population size of the population being examined (see Chapter 7).

Stability Analysis: The effects of variation of the model parameters on LD estimates was examined by replicating the simulation with varied parameters. Three areas need to be examined for their effect on the final LD estimates; the value of N_e used in the construction of the initial haplotypes, the parameter estimates in the models for population dynamics and the model of male breeding success. The effect of the initial value of N_e is examined by both reducing and increasing the assumed value of N_e by a factor of two. For the examination of the effects of population model parameter estimates, an approximate 95% confidence interval was constructed for each coefficient of residual population size. These

were then categorized into groups of high and low effect of residual population size. The simulation was run using these groupings representing the extremes of possible variation caused through the modelling population dynamics. The effect of the estimated distribution of male breeding structure was examined by simulating with the coefficient of variation of the assigned success variable reduced and increased by a factor of two.

6.3 Results

Figure 6.1 plots the simulated LD against distance for 1000 locus pairs. As expected from population genetic theory, LD was significantly negatively correlated with marker distance ($r = -0.4188$, $p < 0.0001$). There is a large amount of variation in the amount of linkage disequilibrium between loci separated by a small distance. As with the average LD value, this decreases as the distance marker pairs are separated by increases, with no absolute D' value over 0.2 being observed for marker pairs separated by greater than 15cM. As markers separated by greater than 20cM showed no significant relationship with distance ($p = 0.444$), these can be use to estimate non-syntenic or background LD. These 490 marker pairs have a mean D' statistic of 0.026 and standard deviation 0.026. This indicates low levels of LD between non-syntenic marker pairs. Among marker pairs separated by less than 10cM, 79% (219/276) have D' greater than 0.026 and 22% (62/276) have D' greater than 0.2. The corresponding LD structure obtained with the use of the r^2 statistic is given in Figure 6.2 and shows a similar pattern to overall LD values.

The stability of the LD estimates with respect to variation in the parameters used to simulate the population is examined in Figure 6.3. The main effect of variation of the parameters used in the simulation is the changing of the average amount of LD at closely linked loci. The distance over which the linkage

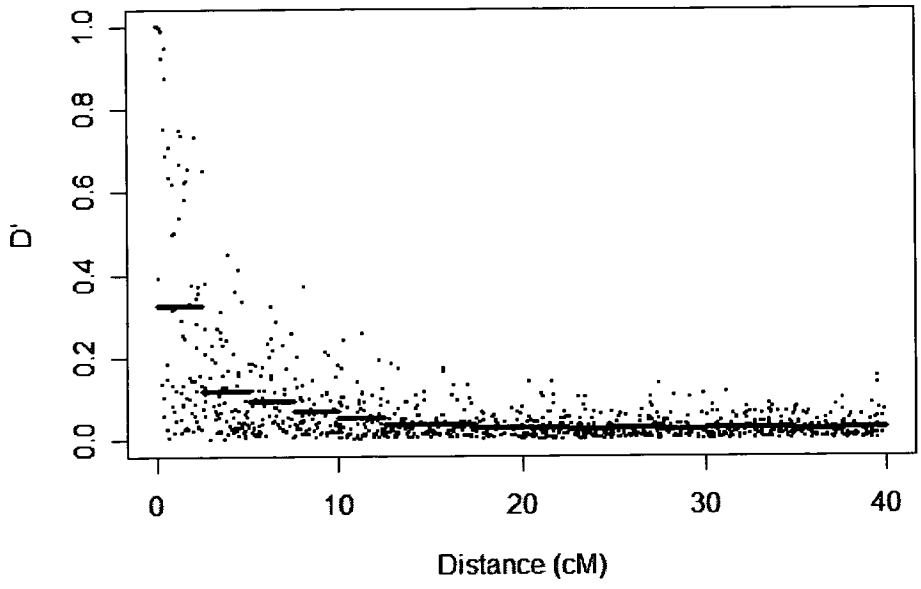

Figure 6.1: Simulated linkage disequilibrium structure of the Soay sheep population. Each point is the absolute value of the D' statistic from a random spaced pair of loci from one simulation replicate (1000 total). The mean D' value over successive intervals of 2.5cM is given by a horizontal line.

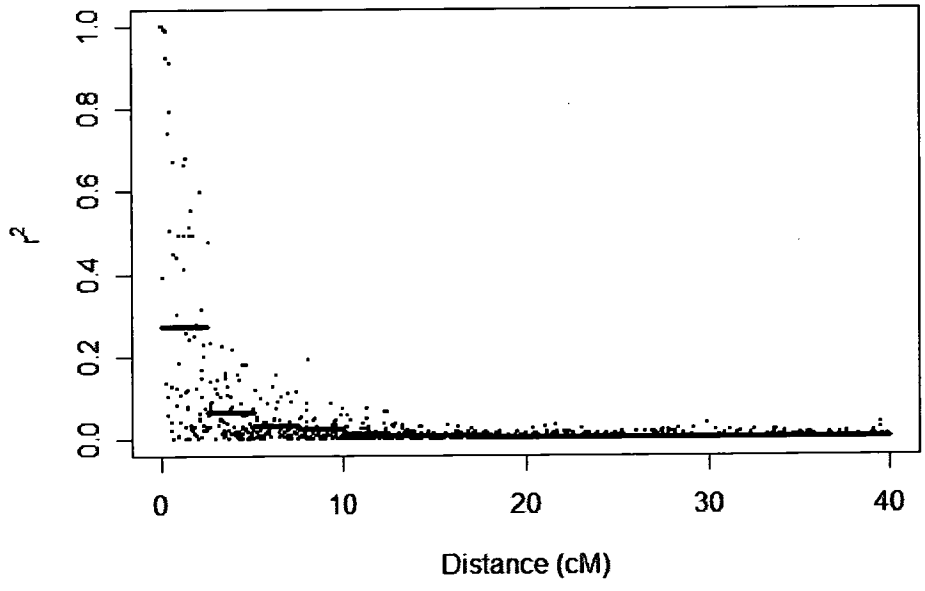

Figure 6.2: Linkage disequilibrium structure as given in Figure 6.1 but measured using the r^2 statistic.

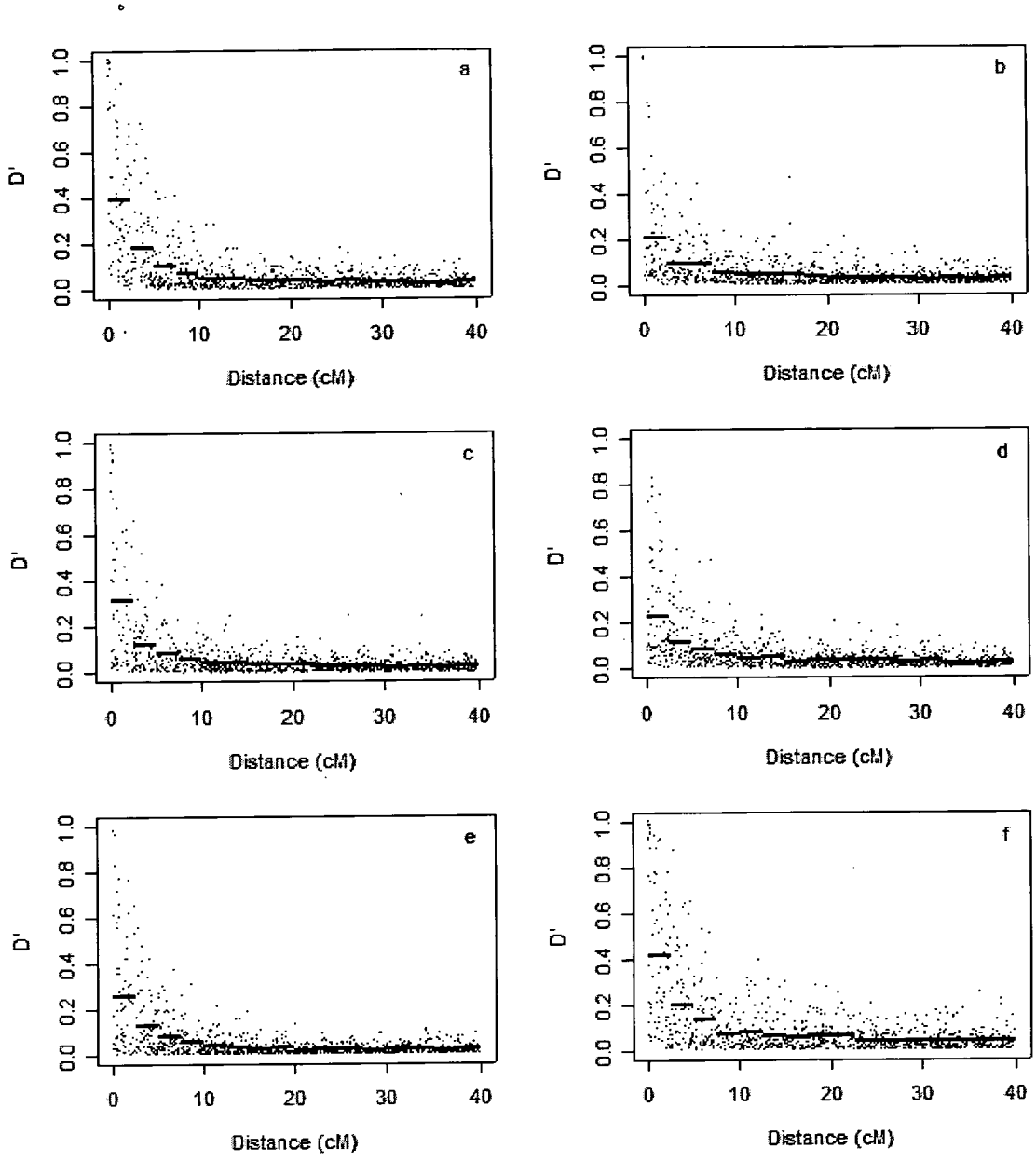

Figure 6.3: Analysis of the stability of simulated linkage disequilibrium structure to variation in population parameters. Each plot shows the linkage disequilibrium structure as given by 1000 simulated replicates using altered parameter values: a) decreasing the effective population size used to simulation the initial haplotypes by a factor of two, b) increasing the initial effective population size by a factor of two, c) using population dynamics parameters at the 'low' end of their 95% confidence interval, d) population dynamics parameters at the 'high' end of their confidence interval, e) decreasing the coefficient of variation of male reproductive success by a factor of two, f) increasing the variation of male reproductive success by a factor of two.

disequilibrium decays and the amount of LD at more distantly linked loci remains similar with all parameter values. The variation of the estimated population dynamics parameters has the least effect on the LD structure with altering the initial effective population size and variation of male breeding success having similar effects.

6.4 Discussion

The simulated Soay sheep population showed considerable amounts of LD between tightly linked loci. As expected there is a significant decline in the amount of LD with distance. Perhaps the most striking feature about this decline is the low level of LD between effectively unlinked marker pairs. Previous studies of LD in domestic livestock have observed a mean non-syntenic LD of 0.211 in Coopworth sheep (McRae *et al.*, 2002), 0.12 to 0.20 in Dutch black and white dairy cattle (Farnir *et al.*, 2000), 0.39 in the UK dairy cattle population (Tenesa *et al.*, 2003) and between 0.11 and 0.22 in domestic pig lines (Nsengimana *et al.*, 2004). The simulated Soay sheep population gave a mean D' value of 0.026 for markers separated by more than 20cM. McRae *et al.* (2002) showed that using a small number of haplotypes inflated the value of D'. Here a total of 3778 haplotypes were used in the estimation of D', resulting in a very low inflation of D' compared to Farnir *et al.* (2000) who used 581 to 1254 haplotypes, McRae *et al.* (2002), 276 haplotypes, Tenesa *et al.* (2003), $<$ 100 haplotypes, and Nsengimana *et al.* (2004), 184 to 302 haplotypes. Correcting the mean values based on the model given by McRae *et al.* (2002) accounts for most of the variation in the mean values. However, the infered range of LD for marker pairs separated by greater than 60cM is lower in the simulated Soay sheep population giving a maximum of 0.2, which is under half of the value given by non-syntenic markers in domestic livestock studies (Farnir *et al.*, 2000; McRae *et al.*, 2002; Tenesa *et al.*, 2003). In this study, marker genotypes were simulated

to represent SNP loci and restricted to have minor allele frequency of greater than 0.1. This is likely to account for some of the differences observed in LD structure as the previous studies have used microsatellite loci and Hedrick's (1987) multiallele D' extension which is susceptible to being biased by small allele frequencies. A further explanation for the differences in the patterns of background linkage disequilibrium observed between domestic livestock populations and the simulated Soay sheep population is concurrent selection at unlinked loci. This will increase the range of linkage disequilibrium at unlinked loci without dramatically increasing mean levels and will not be observed in the simulated population as selection was not included in the population models. However, Farnir *et al.* (2000) found no significant evidence for selection causing an increase in LD between unlinked loci and showed drift accounted for most of the variation in linkage disequilibrium observed in the Dutch black and white dairy cattle population. Thus, the cause of the difference between background LD observed in domestic livestock and the simulated Soay sheep population requires further investigation, possibly with direct measurement of LD in the Soay sheep population.

From the structure of LD given by the simulated Soay sheep an estimate of the required marker density for LD mapping can be obtained. Considering regions with consecutive LD statistics that are overall significantly greater than the mean LD statistic for non-linked markers and single LD statistic outside the range of LD for non-linked markers as indicators of linkage, a minimum marker density for LD mapping in the Soay sheep of approximately 2cM can be obtained. The latest linkage map of the sheep genome has a marker density on average of 3.4cM (Maddox *et al.*, 2001). Slate *et al.* (1998) found only 42.5% of the bovine markers that amplified in the Soay sheep were polymorphic, indicating a higher density linkage map will be required. However, this density is likely to be achievable

with the use of SNP markers. This indicates that the Soay sheep population is a viable resource for linkage disequilibrium fine mapping of quantitative trait loci.

One potential pitfall of using a simulation study to assess the LD structure of a population is an incorrect specification of population dynamics and history. As the demography of the historical Soay sheep population is not recorded, a coalescent model was used to simulate the initial chromosomes. This requires an estimate of effective population size that was derived from the heterozygosity observed in the current population. From the analysis of the effects of misspecification of this parameter, it can be seen that changing this value has little effect on the overall conclusions of this study; LD still decays rapidly from its initial value and the amount of background LD is small. The main difference is the average LD between markers separated by small distances, which increases when N_e decreases and *vice versa*. A decrease in the average LD value at small distances causes the required density of loci needed to fine-map QTL to increase. However, given the estimated ratio of effective population size to census size estimated in Chapter 6, a value of N_e double the estimate (as used in Figure 6.3b) is unlikely. Similarly, the parameter values used to test the effect of the estimation of population dynamics models and variation of male breeding success are extreme values. Even using these, the overall LD structure does not deviate extremely for that observed using the average parameter values, indicating that the estimated LD structure provides an accurate representation of the actual LD structure of the Soay sheep population.

7 Estimating the Effective Population Size of the Soay Sheep Population

7.1 Introduction

In finite populations, the frequency of a particular (non-fixed) allele will vary from generation to generation as a result of genetic drift. In the simplest of cases, the Wright-Fisher idealised population of size N, the variation in allele frequency is a result of a random sampling process in which the alleles passed on to the next generation are distributed as a multinomial process, with sample size $2N$ and allele frequencies p_i as given by the previous generation (Fisher, 1930; Wright, 1931). In this case the variance in allele frequency between generations is given by

$$\sigma^2_{\Delta p_i} = \frac{p_i(1-p_i)}{2N}. \tag{7.1}$$

In large populations, this variation may be negligible but genetic drift rapidly becomes an important evolutionary process as N decreases. Any deviation from random sampling of alleles for the next generation (including unequal contribution from individuals, variable population size, non-random mating and overlapping generations) will affect the amount of genetic drift observed between generations. The amount of drift in an non-idealised population is quantified by the effective population size, N_e, which is the size of the idealised population subject to the same amount of drift.

Under the conditions of the idealised population, the variance in allele frequency in generation $t-1$ is related to the amount of inbreeding in generation t. The coefficient of inbreeding, F, is defined as the probability that the two gametes that combine to form a zygote are identical by descent (Wright, 1922) and is

related to the population size of an ideal population as

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}. \tag{7.2}$$

From this, the proportional increase in the average inbreeding coefficient of an ideal population per generation can be calculated as

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} = \frac{1}{2N}. \tag{7.3}$$

and by comparison with Equation 7.1, the relationship between drift and inbreeding is readily obtained. Although this relationship does not hold in all cases, the estimates for effective population sizes based on variance of allele frequency change and inbreeding (referred to as variance and inbreeding effective sizes, respectively) are usually similar and in practice their sampling variance is large compared to their expected difference.

Despite being a somewhat abstract property of a population, effective population size has become an important parameter in conservation biology. The primary use of effective population size is as an indicator of the long-term survival potential of the population (Soule, 1986; Frankham, 1995b; Schwartz *et al.*, 1998). Although the exact relationship between N_e and survival is debatable, its relationship to molecular evolution is better defined. The amount of drift in a population affects the ability of selection to remove selectively disadvantageous genes (Kimura, 1983). Furthermore, genetic drift removes neutral variation in the genome that may be needed for the adaptation to subsequent changes in the environment. Formally, the relationship between population size and heterozygosity in an ideal population is given by

$$\frac{H_t}{H_{t-1}} = 1 - \Delta F = 1 - \frac{1}{2N}. \tag{7.4}$$

Thus, the effective population size of a population can be estimated through the monitoring of the heterozygosity at a neutral locus (or, more accurately, multiple loci) over time. The effective population size is also related to the amount of linkage disequilibrium between loci. For a population with constant effective population size, the expected value of r^2 measured at two loci separated by recombination fraction c is given by

$$E(r^2) = \mathrm{Var}(\sqrt{r^2}) = \frac{1 - 1/n}{1 + 4N_e c} + \frac{1}{n} \tag{7.5}$$

where n is the sample size collected from the population and $E(r) = E(D) = 0$ (Hill and Weir, 1994). For populations with effective population sizes changing linearly, N_e is replaced with N_{et}, the effective population size $t = 1/2c$ generations ago (Hayes *et al.*, 2003).

Given the theoretical nature of the concept of effective population size and the difficulties associated with accurately estimating the change in heterozygosity in many populations of interest, a variety of predictive equations have been developed for the estimation of effective population sizes in non-ideal populations (reviewed in Caballero, 1994; Wang and Caballero, 1999). However, these are typically limited by the various assumptions invoked to make the solution mathematically tractable and as such the application of the formulae to natural populations is to be undertaken with caution.

An important development in the estimation of effective population size is the extension from discrete to overlapping generations. This started with the theoretical case of a haploid population of size N that randomly loses one individual per time unit (Moran, 1962). Although unrealistic due to the lack of ageing in the model, this case emphasised the decoupling of time units and

generation time (which was defined to be N time units). The estimation of effective population size in populations with overlapping generations has been extended to account for age structure (Felsenstein, 1971; Johnson, 1977; Emigh and Pollack, 1979) and further to account for variation in lifetime family sizes (Hill, 1972a; Hill, 1979). However, these extensions still assume a constant population size, a stable age structure and no heritable component to the variation in lifetime breeding success. Hill (1972b) proposed the concept of annual effective population size, N_y, as a means to compare the effective population sizes of populations with different generation intervals. The annual effective population size is defined as the the size of an idealized population (with generation interval of one year) that has the same variance of genetic drift as the population of interest. The annual effective population size is calculated from the effective population size through the relationship

$$N_y = N_e L \tag{7.6}$$

where L is the mean age of the parents when their progeny are born.

The fluctuations in the population size of the Soay sheep raises questions about the evolutionary consequences of such instability (Clutton-Brock *et al.*, 2004a). Knowledge of the ratio of effective population size to census size is essential for the understanding of such problems. For example, with a low ratio of effective to census size ratio, the fluctuations in population size are essentially buffered with respect to selective forces as the fluctuations in effective population size will have a smaller absolute magnitude. In this chapter, the effective population size of the Soay sheep population is estimated through the use of the simulation model developed in Chapter 6. This is achieved both through monitoring of the amount of inbreeding during the simulation and by using a predictive equation with population parameters as generated in the simulation.

7.2 Methods

Simulation Model: The overall structure of the simulation model of the Soay sheep population is described in Chapter 6. The generation of the initial haplotypes used for modelling linkage disequilibrium structure is not needed for the estimation of effective population size as only the relative change of the inbreeding coefficient or heterozygosity in the population is used. Instead, the initial alleles at the simulated locus were modelled as being uniquely identifiable across the whole population as this set-up allows for increased accuracy of the estimates of the proportional decrease in heterozygosity.

Estimation of Effective Population Size: The estimation of the effective population size of the Soay sheep population was achieved using both the reduction in heterozygosity levels observed over time in the simulated population and through the use of a predictive equation with the generated population dynamics.

For the approach using the simulated heterozygosity values, the annual effective population size for each year of the simulated population was estimated by

$$N_{y(i)} = \frac{1}{2E\left(\Delta F_{(i)}\right)} \tag{7.7}$$

where $E\left(\Delta F_{(i)}\right)$ is the average proportional increase in the inbreeding coefficient (see Equation 7.3) calculated as

$$E\left(\Delta F_{(i)}\right) = 1 - E\left(\frac{H_{(i)}}{H_{(i-1)}}\right) \tag{7.8}$$

with $E\left(H_{(i)}/H_{(i-1)}\right)$ being estimated by average value of $H_{(i)}/H_{(i-1)}$ as obtained from the simulation model. The effective population size is obtained using Equation 7.6 with $L_{(i)}$ being given by the mean age at the time of lambing of

all sheep that were available to contribute offspring at that lambing period, independent of their survival to the lambing period. All values were estimated by their mean over 10000 replicates.

Approximate 95% confidence intervals for the estimates of effective population size are derived by noting that unless N_e is very small, the expected value of the difference between H_t and H_{t-1} is small and thus the proportional change in heterozygosity will be close to zero. As this value occurs in the denominator of Equation 7.7, it will be the source of the majority of the variance in the estimates of N_e. By the central limit theorem,

$$\Delta F_{(i)} \xrightarrow{D} N\left(\mu_{\Delta F_{(i)}}, \frac{\sigma^2_{\Delta F_{(i)}}}{n}\right) \tag{7.9}$$

where the mean and variance of $\Delta F_{(i)}$ can be calculated through its relationship to $H_{(i)}/H_{(i-1)}$ and n is the number of simulation replicates. Assuming the variation in the estimation of $L_{(i)}$ is relatively small, approximate 95% confidence intervals for $N_{e(i)}$ can be constructed by inputting the confidence limits for $\Delta F_{(i)}$ in Equation 7.7.

Estimation of the effective population size of the Soay sheep was also undertaken using a general predictive equation. The equation developed by Hill (1972a; 1979) is applicable to the Soay sheep population as it accounts for unequal sex ratio, variation in reproductive success and overlapping generations. However, the equation requires knowledge of covariances that are not readily estimated in practice. Although the population dynamics in this case are generated by simulation and thus these covariances are estimable, it is of greater interest to examine the properties of an estimator that can be more readily applied to actual population data. The predictive equation has been simplified by Nunney (1991;

1993) who removed these covariances by assuming the number of adults of both sexes is much greater than the generation time. This gave the estimate of effective population size for a population of size N with a sex ratio of r males to $1-r$ females as

$$N_e = \frac{4r(1-r)NT}{\left(\begin{array}{c} (A_m(1-r)+A_f r)-(2r/b_f)+ \\ (I_{bm}(1-r)+I_{bf}r)+(A_m I_{Am}(1-r)+A_f I_{Af} r) \end{array} \right)} \tag{7.10}$$

where T is the mean age of all individuals at breeding (in breeding seasons), A_m and A_f are the average lifespan for males and females respectively and b_m and b_f are the average fecundity per breeding season. The remaining parameters, in the form of I_p, are the standardized variances for parameter p calculated as the variance divided by the mean squared (i.e. the coefficient of variation squared). The average lifespan of the Soay sheep was estimated using the equilibrium age structure that was also used to construct the ages of the initial sheep in the simulation. This gave values of $A_m = 3.15$, $A_f = 3.34$, $I_m = 0.68$ and $I_{Af} = 1.43$. The value of A_m is inflated to account for males not being required to live until the birth of their offspring. Confidence intervals for the estimates of effective population size were constructed from the 0.025 and 0.975 quantiles of the distribution of estimates across simulation replicates.

Stability Analysis: As with the linkage disequilibrium study in Chapter 6, the stability of the estimates of effective population size to the variation of parameters was investigated. The effects of variation in the estimates of linear model parameters and variation of male breeding success were investigated.

7.3 Results

Figure 7.1 shows the average value of the inbreeding coefficient over time assuming that the initial sheep that were transferred to Hirta were unrelated. The rate of increase is at its highest during the initial growth of the population size and decreases to an almost constant rate after ten years. At the end of the simulation the average inbreeding coefficient in the population is 0.071. Through the relationship between the inbreeding coefficient and heterozygosity given in Equation 7.4, this gives an average decrease of heterozygosity of 7.1% as used in the construction of haplotypes for the modelling of linkage disequilibrium (see Chapter 6).

The average age of breeding sheep in the simulation model is given in Figure 7.2. This is calculated as $(L_m + L_f)/2$, where L_m and L_f are the average ages of male and female sheep that were assigned offspring. For the purposes of estimating N_e, these values are calculated at the time of lambing and are independent of survival to that time. This is particularly important with respect to male sheep that can have offspring born after their death. As expected, in the years after the sheep were introduced to Hirta there is a drop in the average age as the population rapidly grows through the birth of new lambs. Average age then rises to an equilibrium level which the population fluctuates about. The initial average age is higher than the equilibrium value despite adult ages being sampled from the equilibrium age. This is caused by the unusual population structure of the sheep that were transferred to Hirta from Soay with male lambs castrated and therefore not forming part of the genetic founder population.

The estimates of effective population size based on the changes in the average inbreeding coefficient with time are give in Figure 7.3. In the initial years of the simulation, the effective population size increases rapidly as the population

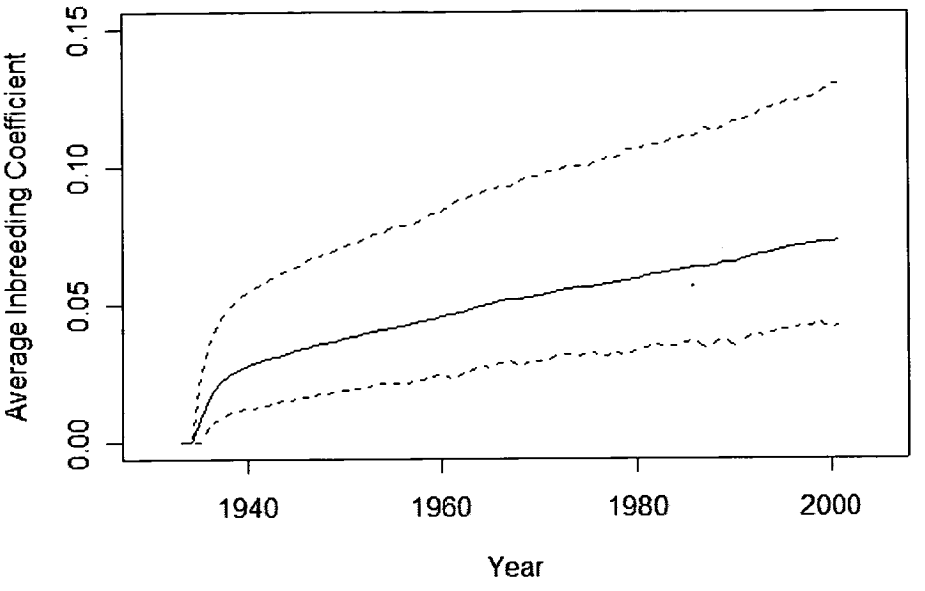

Figure 7.1: The average value of inbreeding coefficient derived from heterozygosity change in the simulated populations assuming founding individuals were unrelated. Dashed lines indicate the symmetric 95% confidence limits of values observed across simulation replicates.

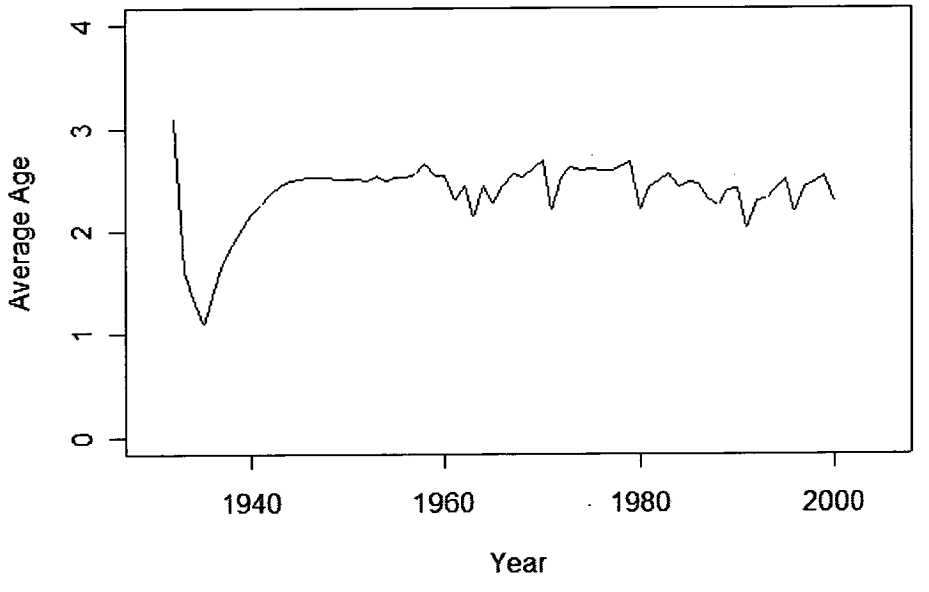

Figure 7.2: Average age at the time of lambing for sheep assigned offspring. For male parents this is independent of survival to the lambing period.

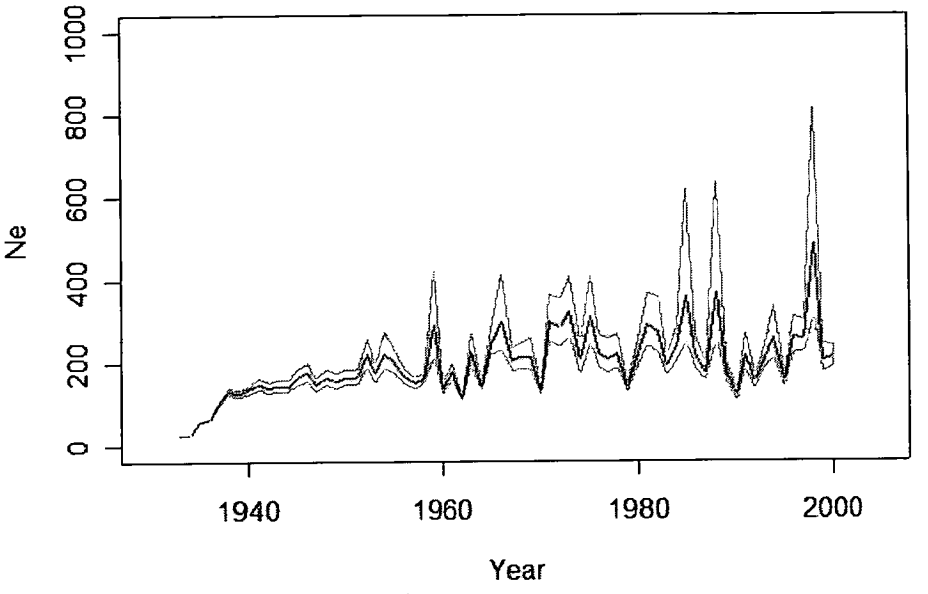

Figure 7.3: Estimated value of effective population size using the average change in average inbreeding coefficient with time. The shaded area represents approximate 95% confidence intervals.

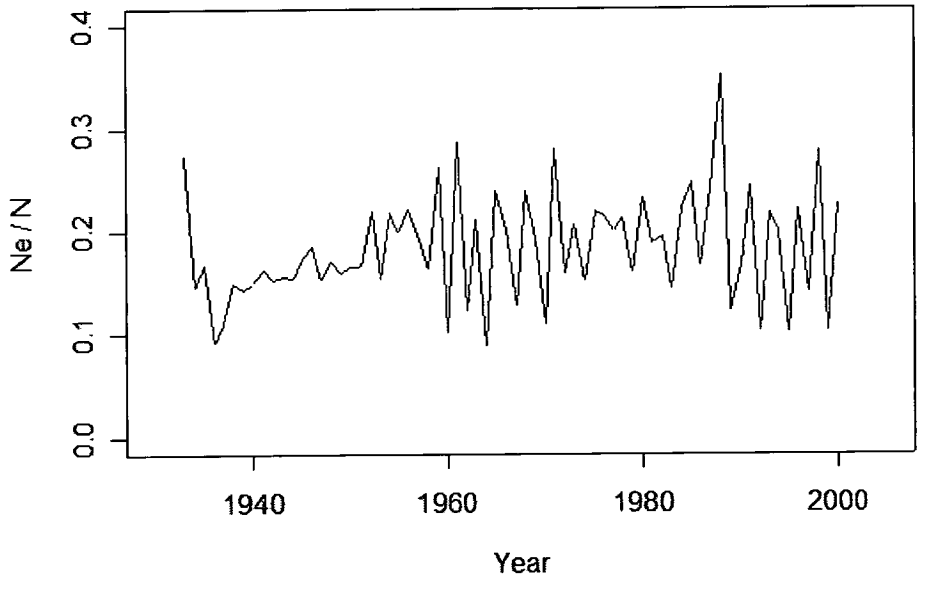

Figure 7.4: Ratio of effective to census population size. After the initial period of population growth where the ratio drops rapidly then stabilises. The ratio fluctuates with harmonic mean of 0.17 during the period where actual census data is known.

becomes established. This is followed by a period when the effective population size shows a slight increase with time and shows little fluctuation about the trend. This relative stability is due to the averaging of the effects of fluctuating population size over the replicates. In the final stage of the simulation, the fluctuations in population census size cause similar fluctuations in the yearly estimates of effective population size. These deviations in the values of effective population size from the overall trend are much smaller than those of census size. Figure 7.4 shows the ratio of effective population size to census size during the simulation. Using the period of the simulation during which actual population census sizes are known, the harmonic mean of this ratio is 0.170. This value is consistent with the value of N_e estimated through the use of observed heterozygosity and used in the simulation of haplotypes for the modelling of linkage disequilibrium in Chapter 6.

The values of effective population size estimated through the use of the predictive formula are given in Figure 7.5. These estimates have similar characteristics to those estimated through the monitoring of inbreeding, with an initial increase in N_e followed by a relatively stable period before fluctuating about the trend line. The large confidence intervals for the estimates of effective population size during the period that population size was considered to fluctuate randomly highlight the difference in the conceptual development of the confidence intervals for the two estimates. Because the increase in the amount of inbreeding each year is very small, estimates of effective population size based on one simulation replicate may give an infinite value for a particular year. Thus, confidence intervals for these estimates are based on the statistical variation about the estimated mean increase of inbreeding per year. With a predictive formula, the estimates of effective population size are much less variable allowing confidence intervals to represent variation across replicates instead.

The estimates of effective population size using the predictive equation are approximately twice as large on average as those estimated using the amount of inbreeding (Figure 7.6). From the ratio of the effective size estimates during the initial stages of the simulation, population growth appears to be a major contributor to this difference. One likely source of this difference is the use of a constant average lifespan in the denominator of the predictive equation. In years when the population size is increasing, the average lifespan of individuals in the Soay sheep population is known to be longer than those born before a population decline. This effect would have been especially large in the initial colonization of Hirta where old age would have been the primary cause of death. Although adjustments for such effects can be achieved when using a simulation model or through the use of long term field data, most applications of the predictive equation would require the average lifespan be estimated using only a few years of data. Also, not adjusting for this effect allows the examination of the robustness of the estimator to deviations from the assumed population structure, which is the primary reason for including the second estimator in this study.

Figure 7.7 examines the effects of variation of model parameters on the estimates of effective population size. The harmonic mean of the effective population size estimates after 1960 is used for the purpose of comparing the different models as the yearly estimates of effective population size show a reasonably flat trend after this time. The main simulation model gives an estimate of average effective population size of 205 over this period. The variation in the estimates of the statistical models for sex ratio and survival have negligible effect on the estimated effective population size (Figure 7.7a and b), giving average effective population sizes of 206 and 196 for the models where population size has a lesser or greater

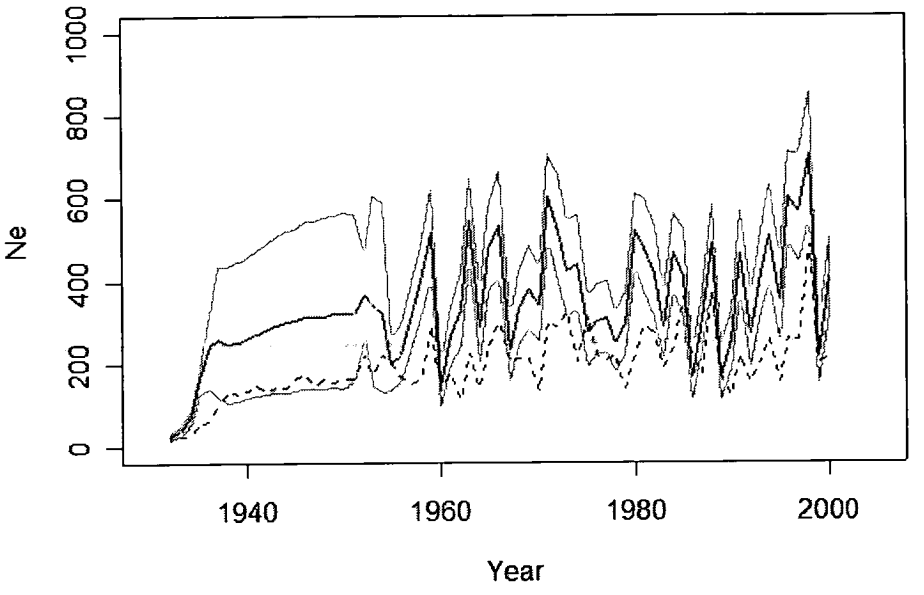

Figure 7.5: Effective population size estimates obtained using a predictive equation. The shaded area represents the 95% confidence interval of the estimate over simulation replicates. The estimates obtained by the change in average inbreeding are provided for comparison (dashed line).

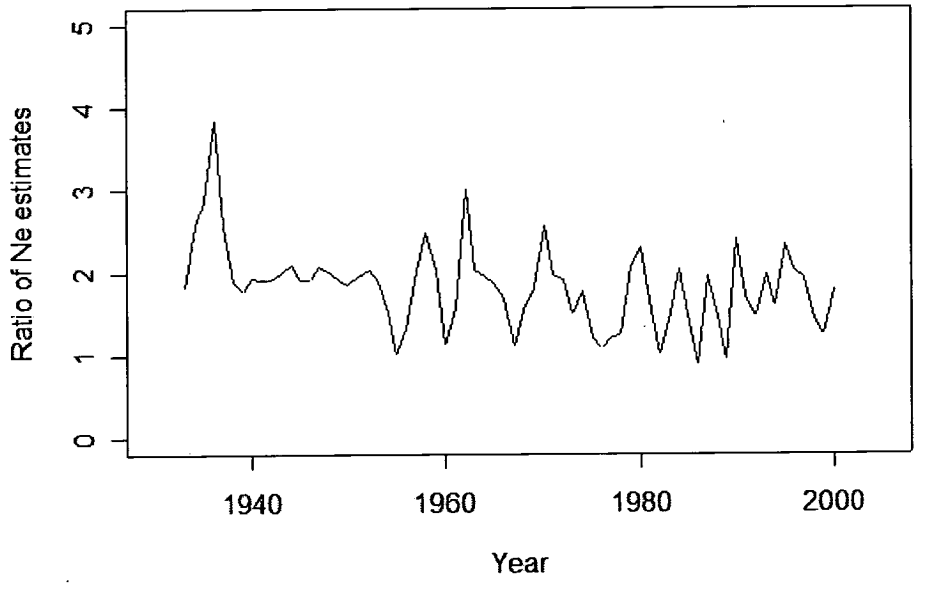

Figure 7.6: Ratio of the estimate of effective population size calculated using the predictive equation to those obtained through the increase in the average amount of inbreeding.

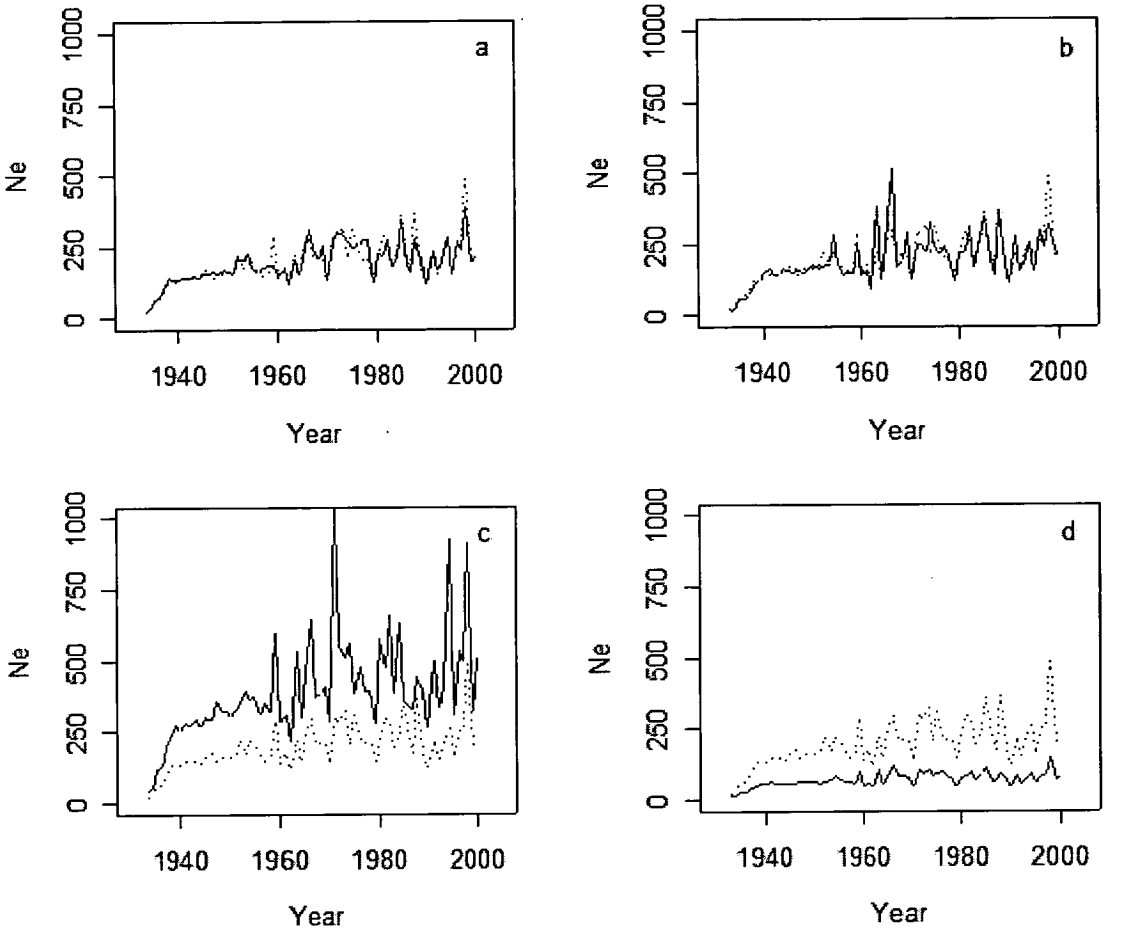

Figure 7.7: Analysis of the stability of simulated effective population size estimates to variation in population parameters. Each plot shows the average effective population size estimate from the decay in heterozygosity over 10000 simulation replicates using altered parameter values: a) using population dynamics parameters at the 'low' end of their 95% confidence interval, b) population dynamics parameters at the 'high' end of their confidence interval, c) decreasing the coefficient of variation of male reproductive success by a factor of two, d) increasing the variation of male reproductive success by a factor of two. The dashed line indicates the estimated effective population size from the initial model (see Figure 7.3).

effect on population dynamics respectively. In contrast, altering of variation in male reproductive success has a large effect on the estimates of effective population size (Figure 7.7c and d) giving average estimates of 404 and 72 for the cases where the variation in male reproductive success was reduced and increased by a factor of two.

7.4 Discussion

Estimation of the effective population size of the Soay sheep population allows the examination of the effect of the large fluctuations of population size on an evolutionary scale. From a comparison of Figures 5.2 and 7.3, the absolute magnitude of the fluctuations in effective population size is clearly smaller than those of census size. However, the coefficient of variation of the fluctuations after 1955 is larger for effective population size (0.33) than for census size (0.28), although this difference is not significant (bootstrap p-value = 0.27). This indicates that although the ratio of effective population size to census population size varies over time, simplifying any evolutionary model of the Soay sheep to have a constant ratio of $N_e = 0.17N$ will not adversely alter any results obtained.

The stability of estimates of effective population size with changing parameter values indicates that assumed variation in male breeding success is the primary source of uncertainty in the simulation results. In order to gain a better understanding of the effect of variation in male reproductive success, Equation 7.10 can be reduced to only include the standardized variance of male reproductive success (I_{bm}) giving

$$N_e = \frac{a}{bI_{bm} + c} \tag{7.11}$$

where parameters a, b and c can be readily inferred by comparison of the equations. This clearly shows that an increase in the variance of male breeding

success will decrease the estimates of effective population size, and vice versa, but these changes will not be linear due to the addition of c in the denominator. The stability of the simulated estimates of effective population size was investigated by increasing and decreasing the coefficient of variation of the male breeding success by a factor of two, being equivalent to increasing and decreasing I_{bm} by a factor of four. Using these values, the harmonic average of effective population size was estimated to be 35% and 197% of the estimate given by the main simulation model. As these estimates are close to the 50% and 200% expected if the value of N_e was considered to be proportional to the inverse of the coefficient of variation of male reproductive success, a linear adjustment of the estimate of effective population size can be used as an approximate correction when improved information on the distribution of male reproductive success becomes available.

Although the simulation model of the Soay sheep population includes the major aspects of the observed population dynamics, other known features of the Soay sheep population will cause deviations from the estimates of effective population size. In the simulations reproductive success was considered not to have a heritable basis. Excluding this heritability in the modelling of female reproduction is likely to have limited effect on the final results due to the relatively uniform distribution of female reproductive success in a single breeding season. However, the assumed lack of heritability in male reproductive success is likely to upwardly bias the effective population size estimates (Wray *et al.*, 1990). Another possible cause of bias in the estimates of effective population size is population substructure. The portion of the Soay sheep population in the Village Bay area has been shown to have substructure by both the clustering of measured distance between pairs of individuals (Coulson *et al.*, 1999) and through the use of molecular markers (Coltman *et al.*, 2003). This grouping

is primarily caused by females adopting similar home-ranges to their mother. However, the male sheep in the Village Bay area disperse throughout the region and there is lack of evidence for departure from Hardy-Weinberg equilibrium at marker loci investigated in whole cohorts (Coltman *et al.*, 2003). This indicates the effect of this substructure on the amount of inbreeding in the population is minimal, although further substructure is likely to exist at different scales on the island.

The disparity between estimates of effective populations size based on (simulated) demographic and molecular data has been previously observed, although not to the scale observed here. Harris and Allendorf (1989) simulated populations with dynamics similar to that of grizzly bears. These simulations included fluctuations in population size with standard deviation of 10% of the average population size. It was shown that estimation of effective population size using Hill's (1972a) equation results in approximately a 5% upward bias in effective population size estimates. However, the extent of the bias was restricted when compared to estimates obtained in this study due to the overall age structure remaining relatively stable throught the simulation.

The observation that even very general predictive equations can be severely biased by deviations from the assumed population structure suggests that effective population size should also be estimated using other methods to provide an independent assessment of the accuracy the predictive equation. The use of molecular methods (through either measuring the change in allele frequency or amount of inbreeding) is hindered by the large variances that accompany the associated estimates of effective population size and the requirement of loci used to be neutral with respect to fitness. Using population data to build simulation models, even when only including the major components of population

structure, may provide the most accurate methodology for the estimation of effective population size in populations where there are large deviations from the assumptions made in predictive equations.

8 General Discussion

8.1 Summary

The primary strategy for the dissection of the genetic architecture of quantitative traits is through the use of linkage analyses that localize regions of the genome containing loci affecting the trait(s) of interest. Once these regions are located, the goal shifts to identifying the underlying genetic polymorphism that affects the trait. This involves the reduction of the size of the linkage region, primarily through information from linkage disequilibrium. In this thesis, linkage mapping was performed on a commercial sheep pedigree. Also, methodology for the evaluation of significance thresholds in linkage mapping was examined. A simulation study was used to investigate the possibility and aid the design of linkage disequilibrium mapping in a feral sheep population.

In Chapter 2, QTL mapping is performed across candidate chromosomes in a large commercial Charollais sheep pedigree. Chromosomes were chosen for investigation using evidence for QTL of large effects from other studies in sheep as well as other livestock species. A total of nine QTL were detected at the 5% chromosome wide significance threshold, two from the analysis of a moderately sized half-sib family within the complete pedigree and seven from a variance component analysis of the whole pedigree. Of these nine QTL, only one mapped to a candidate region. Given none of the novel QTL were highly significant, further confirmation studies are needed before these are targeted for fine-mapping or selection. While the results do not demonstrate a lack of variation at the loci underlying the candidate regions, they demonstrate that what may be considered a locus of major effect in one breed, may not in another due to differences in the distribution of allelic effects between the populations. This indicates the need to examine commercially important traits in a variety of

breeds as this will allow the comparison of the distribution of QTL across the genome and the size of their allelic effects. While this information is important in designing selection schemes in different populations, it would also be useful in the search for the genetic variants underlying QTL. Efforts could be focused on breeds where the effect of this QTL is the largest, thus increasing the power to localize the variant.

Chapter 3 examines a region of sheep chromosome 1 showing a difference in marker order across breeds. Strong statistical evidence is given for the existence of three different linkage map orders across four sheep populations. Ideally, these differences require confirmation, which may be achieved either statistically through the use of an independent mapping panel or physically with the use of fluorescent *in situ* hybridisation. The most parsimonious solution to the history of this chromosomal region in these requires two inversions. Considering the time since the divergence of these breeds, this indicates that this chromosomal region is relatively unstable. The number of such regions in the sheep genome requires further examination. However, this does indicate that some caution is necessary when implementing QTL mapping experiments involving the crossing of distantly related breeds as underlying assumptions about the nature of the underlying linkage map may be violated.

Methodology for the evaluation of significance thresholds for QTL mapping using variance component methodology on data from general pedigrees is investigated in Chapter 4. The examined method estimated significance thresholds through the permutation of independent groups within the pedigree that are formed subject to constraints on relationships to other pedigree members. These constraints maintained the trait heritability in the pedigree and thus maintained a constant null likelihood. The significance thresholds estimated using this

method showed a downwards bias as the proportion of the pedigree being permuted was reduced. Upwards bias was seen when QTL of large effect was simulated. The source of this bias is identified as being due to the non-permuted part of the pedigree restricting the exploration of the null distribution. How to randomize this information needs to be the focus of further research in this area.

Chapters 5 to 7 investigate a feral Soay sheep population through the use of a simulation model. In Chapter 5, a unified statistical framework was constructed for the modelling of the population dynamics of the Soay sheep of St. Kilda, Scotland. This population show dramatic fluctuations in population size between years. The statistical models examined the effects of population size on the adult sex ratio, proportion of lambs and survival. These models were used to form a simulation study examining the linkage disequilibrium structure of the population (Chapter 6) and its effective population size (Chapter 7). This provides the groundwork for the design of future fine-mapping studies in this population.

This simulated linkage disequilibrium structure showed considerable amounts of linkage disequilibrium between tightly linked loci and a strong decline in the average amount of linkage disequilibrium with genetic distance. In contrast to studies in domestic livestock, the simulation showed low levels of linkage disequilibrium between markers that were separated by large genetic distances. Although the reasons for this need to be investigated further, the different types of markers used in this simulation study and the published studies in domestic livestock is likely to a be major source. Further differences will be attributable to the contrasting histories of the two populations. The published linkage disequilibrium structure of domestic sheep is primarily estimated in Coopworth sheep, a breed that was developed in the 1960s from the Border Leicester and

Romney breeds. This differs from the Soay sheep population, which has been isolated for many centuries. Thus, the population history of a breed being studied needs to be considered when designing genetic association studies. The predicted linkage disequilibrium structure in the Soay sheep population was relatively stable to variation in the model parameters. From the simulation, a marker density of at least 2cM is recommendable for fine mapping using linkage disequilibrium in this population.

The effective population size of the Soay sheep population is estimated in Chapter 7. This was achieved using two methods; the direct monitoring of inbreeding through the decrease of heterozygosity at a simulated locus and indirectly through the use of a general predictive equation with the population parameters obtained from the simulation. The two methods showed a great disparity between estimated effective population sizes, with the estimates from the predictive equation being approximately twice those from the monitoring of heterozygosity. The bias in the estimate of the predictive equation occurs because of the inadequate description of population dynamics. However, the removal of bias from the use of heterozygosity data comes at price of reduced accuracy in the estimates. The effective population size fluctuates about 17% of the census population size after the initial period of population growth. Examination of the effects of variation of the simulation parameters demonstrated that estimates of effective population size were sensitive to the variance of male breeding success. However, approximate corrections can be made using a simple linear transformation when more accurate values are available.

8.2 Future Directions

Most of the research on quantitative traits in sheep is driven by the prospect of commercial gain through the genetic improvement of flocks. However, the current

understanding about the genetic architecture of commercially important traits is still limited. The underlying polymorphisms in loci effecting quantitative traits have only been identified for genes of such large effect that visual inspection can often determine the genotype. The search for quantitative trait loci with moderate to small effect sizes is progressing slowly in comparison to other livestock species and few detected QTL have been replicated. Also, the accuracy in the localization of these loci tends not to have improved from their initial detection. A scan of the proceedings of animal breeding conferences (e.g. the meetings of the European Association for Animal Production, the Association for the Advancement of Animal Breeding and Genetics, and the World Congress on Genetics Applied to Livestock Production) shows large numbers of preliminary results of QTL scans and the creation of many resource populations for QTL mapping, often through the use of selection lines. This indicates that either the number of published results from QTL mapping experiments will increase rapidly in the near future, or that the complete results of many QTL mapping experiments are remaining unpublished. The latter situation, while somewhat justifiable given the potential commercial gain that could be made from this information, slows the overall progress in the understanding of the genetics of quantitative traits in sheep as many experiments will be unnecessarily repeated, draining the already limited resources of the sheep industry.

Knowledge of the genetic architecture of a quantitative trait will help in the improvement of flocks through an increased accuracy in selection procedures (reviewed in Dekkers and Hospital, 2002; Williams, 2005). The identification of the exact gene in a chromosomal region that affects a trait is not needed for the use of marker assisted selection (MAS). However, it will allow the relatively quick examination of whether polymorphisms in this gene are segregating within other populations. Without this information, QTL mapping experiments will need

to be performed over all populations where this trait is of interest. Also, MAS of large chromosomal regions may have negative effects on other traits through linkage disequilibrium with unfavourable alleles at other genes (Williams, 2005). Given the large amount of variation between sheep breeds compared to within breed variation, it will eventually become desirable to introgress favourable alleles from one breed to another. This can be achieved while maintaining the genetic background of the recipient breed with the use of marker assisted introgression (MAI; Hospital and Charcosset, 1997). The localization of genes to very small chromosomal regions is especially important for MAI as this limits potential negative alleles in surrounding genes hitch-hiking with the selected gene.

Linkage disequilibrium mapping is becoming the primary strategy in the fine mapping of genes to regions of suitable length for candidate gene studies. The potential of linkage disequilibrium to localize genes of interest has not yet been realized in sheep. However, linkage disequilibrium fine-mapping has had several notable successes in cattle and pigs. In cattle, a QTL with large effect on milk yield and fat and protein percentage was mapped to chromosome 14 (Coppieters *et al.*, 1998; Heyen *et al.*, 1999). Using linkage disequilibrium, this QTL was fine mapped to a region of 3cM (Farnir *et al.*, 2002) and subsequently identified as a polymorphism in the *DGAT1* gene (Grisart *et al.*, 2002). In pigs, a QTL with effects on muscle growth, fat deposition and heart size was identified in crosses of the Large White domestic breed with Piétrain breed and European wild boar (Jeon *et al.*, 1999; Nezer *et al.*, 1999). This was fine-mapped to a region of 250kb (Nezer *et al.*, 2003) in which the underlying polymorphism regulating the expression of the *IGF2* gene was discovered (Van Laere *et al.*, 2003). These species have very similar LD structures to that of sheep (Farnir *et al.*, 2000; McRae *et al.*, 2002; Nsengimana *et al.*, 2004), indicating that the future may hold such successes in sheep genetics.

Improvements in technology will allow the use of methods for the dissection of genetic architecture that are currently unavailable and prohibitively expensive for use with sheep. Looking at recent trends in the better funded field of human genetics, there are two main areas where technology is changing the analysis of complex traits. Whole genome association studies are now being applied in the search for genes underlying complex disease (e.g. Hu *et al.*, 2005; Maraganore *et al.*, 2005; Tamiya *et al.*, 2005). However, the LD structure in livestock is substantially different to that of humans. The much larger distances that LD extends across in sheep compared to humans results in a much lower density of markers across the genome being needed. Thus, while 200,000 to 1,000,000 markers may be considered the minimum for whole genome association studies in humans (Carlson *et al.*, 2004), a number in the order of 10,000 would provide more than adequate coverage in livestock. However, this advantage will come at a price. The larger span of linkage disequilibrium in the sheep genome reduces the accuracy with which QTL are localized. Thus, while this approach in humans may localize a genetic variant to a region containing one or at most a few genes, in sheep this region may contain in the order of ten genes. Even if an obvious candidate gene is in the localized region, the amount of work to determine the underlying genetic variant will be substantially greater in sheep than humans.

A second area beginning to impact the study of complex traits in humans is gene expression studies (Lander, 1999; Cobb *et al.*, 2005). The lack of sequence data for sheep implies this technology won't be available for studies in sheep in the very near future. However, with the draft sequence of the cattle genome now in public databases, this technology may soon be applied to livestock. The sequence of the cattle genome is estimated to have around 90% similarity to that of sheep. Along with the relatively few rearrangements between the cattle and

sheep genome, this provides an important platform from which these techniques can be developed in sheep.

While not ready to be applied to sheep, high throughput genotyping is increasing the understanding of quantitative genetic variation. Both theoretical models and data from QTL mapping and selection experiments demonstrate the distribution of allelic effects for genes underlying quantitative traits is positively skewed, with many genes of small effect and few of large effect (Mackay, 2001; Hayes and Goddard, 2001). Currently, the QTL detected in whole genome linkage studies tend to have large effects on the trait being studied. This is a direct consequence of the increased power to detect genes with large effect sizes compared to those of smaller effect. The advances in genotyping methods allow for the power of QTL mapping studies to be increased, both through the dense coverage of the genome and the ease that sample sizes can be increased. Thus, genes with smaller allelic effects are able to be detected. This will result in a larger proportion of the distribution of allelic effects being examined allowing further validation of theoretical models.

The increase in the identified of causal polymorphisms underlying QTL will also elucidate the nature of the mutations underlying quantitative traits, specifically the relative contribution of structural and regulatory mutations. While structural mutations (mutations that alter the gene product) are known to be the primary cause of Mendelian traits, current evidence cautions against extending this to complex traits (Thomas and Kejariwal, 2004). In fact, the number of regulatory mutations (mutations that alter the expression levels of a gene product) that have been identified or implicated in the expression of quantitative traits is rapidly increasing (e.g. Van Laere *et al.*, 2003; Schnabel *et al.*, 2005; Oliver *et al.*, 2005). While it is still too early to conclusively state

the proportion of QTL caused by regulatory mutations, it is variation due to this type of mutation that is being examined in gene expression studies. Thus, until this is further clarified it will be necessary to continue with the more traditional linkage/association studies that also detect variation due to structural mutations.

The major stumbling block in the future of sheep genetics is a lack of funding. This is aptly demonstrated in a recent review on domestic animal genomics by Andersson and Georges (2004) who tallied the number of expressed sequence tags (ESTs) in the GenBank database at the start of 2004. Sheep had approximately 7000 ESTs, less that 3% of the number for cattle, pigs and chickens and around 50% of the amount for horses. More recent summaries of the GenBank database show this disparity is growing larger. The cost of developing the technology discussed above for use with sheep is currently prohibitively expensive. While relying on results from QTL mapping of meat and milk traits cattle could be used as a low cost alternative to directly studying these traits in sheep, traits such as wool have no such luxury. However, the ever decreasing price of the high throughput technologies will eventually remove many of todays limits.

Bibliography

Aalen, O. (1978). Nonparametric inferences for a family of counting processes. *Annals of Statistics* **6**: 701–726.

Almasy, L. and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**: 1198–1211.

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics* **54**: 535–543.

Andersson, L. and Georges, M. (2004). Domestic-animal genomics: Deciphering the genetics of complex traits. *Nature Reviews Genetics* **5**: 202–212.

Archibald, A. L., Haley, C. S., Brown, J. F., Couperwhite, S., McQueen, H. A., Nicholson, D., Coppieters, W., Van de Weghe, A., Stratil, A., Winterø, A. K., Fredholm, M., Larsen, N. J., Nielsen, V. H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J. C., Burgaud, G., Ollivier, L., Bidanel, J.-P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E. J. M., Groenen, M. A. M., Davies, W., Høyheim, B., Keiserud, A., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J. R., Anderson Dear, D. V., Signer, E., Jeffreys, A. J., Moran, C., Le Tissier, P., Muladno, M., Rothschild, M. F., Tuggle, C. K., Vaske, D., Helm, J. M., Liu, H.-C., Rahman, A., Yu, T.-P., Larson, R. G. and Schmitz, C. B. (1995). The PiGMaP consortium linkage map of the pig (*Sus scrofa*). *Mammalian Genome* **6**: 157–175.

Ardlie, K. G., Kruglyak, L. and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**: 299–309.

Arnold, H., Della-Fera, M. A. and Baile, C. A. (2001). Review of myostatin history, physiology and applications. *International Archives of Bioscience* pages 1014–1022.

Bacanu, S.-A. (2005). Robust estimation of critical values for genome scans to detect linkage. *Genetic Epidemiology* **28**: 24–32.

Band, M. R., Larson, J. H., Rebeiz, M., Green, C. A., Heyen, D. W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J. E. and Lewin, H. A. (2000). An ordered comparative map of the cattle and human genomes. *Genome Research* **10**: 1359–1368.

Barendse, W., Armitage, S. M., Kossarek, L. M., Shalom, A., Kirkpatrick, B. W., Ryan, A. M., Clayton, D., Li, L., Neibergs, H. L., Zhang, N., Grosse, W. M., Weiss, J., Creighton, P., McCarthy, F., Ron, M., Teale, A. J., Fries, R., McGraw, R. A., Moore, S. S., Georges, M., Soller, M., Womack, J. E. and Hetzel, D. J. S. (1994). A genetic linkage map of the bovine genome. *Nature Genetics* **6**: 227–235.

Barendse, W., Vaiman, D., Kemp, S. J., Sugimoto, Y., Armitage, S. M., Williams, J. L., Sun, H. S., Eggen, A., Agaba, M., Aleyasin, S. A., Band, M., Bishop, M. D., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B., Drinkwater, R. D., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harlizius, B., Hawkins, G., Hetzel, J., Hirano, T., Hulme, D., Jorgensen, C., Kessler, M., Kikpatrick, B. W., Konfortov, B., Kostia, S., Kuhn, C., Lenstra, J. A., Leveziel, H., Lewin, H. A., Leyhe, B., Lil, L., Martin Burriel, I., McGraw, R. A., Miller, J. R., Moody, D. E., Moore, S. S., Nakane, S., Nijman, I. J., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Teale, A. J., Thieven, U., Urquhart, B. G. D., Vage, D.-I., Van de Weghe, A., Varvio, S., Velmala, R., Vilkki, J., Weikard, R., Woodside, C., Womack, J. E., Zanotti, M. and Zaragoza, P. (1997). A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* **8**: 21–28.

Baret, P. V., Knott, S. A. and Visscher, P. M. (1998). On the use of linear regression and maximum likelihood for QTL mapping in half-sib designs. *Genetical Research* **72**: 149–158.

Barillet, F., Arranz, J.-J. and Carta, A. (2005). Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genetics Selection Evolution* **37 (Suppl. 1)**: S109–S123.

Barton, N. H. and Keightly, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**: 11–21.

Baylis, M. and Goldmann, W. (2004). The genetics of scrapie in sheep and goats. *Current Molecular Medicine* **4**: 385–396.

Beh, K. J., Hulme, D. J., Callaghan, M. J., Leish, Z., Lenane, I., Windon, R. G. and Maddox, J. F. (2002). A genome scan for quantitative trait loci affecting resistance to *Trichostrongylus colubriformis* in sheep. *Animal Genetics* **33**: 97–106.

Beh, K. J. and Maddox, J. F. (1996). Prospects for development of genetic markers for resistance to gastrointestinal parasite infection in sheep. *International Journal for Parasitology* **26**: 879–897.

Berghmans, S., Segers, K., Shay, T., Georges, M., Cockett, N. and Charlier, C. (2001). Breakpoint mapping positions the callipyge gene within a 450-kilobase chromosome segment containing the *DLK1* and *GTL2* genes. *Mammalian Genome* **12**: 183–185.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L. F., Hawkins, G. A., Solinas Toldo, S., Fries, R., Grosz, M. D., Yoo, J. and Beattie, C. W. (1994). A genetic linkage map for cattle. *Genetics* **136**: 619–639.

Boehnke, M. (2000). A look at linkage disequilibrium. *Nature Genetics* **25**: 246–247.

Bot, J., Karlsson, L. J. E., Greef, J. and Witt, C. (2004). Association of the MHC with production traits in Merino ewes. *Livestock Production Science* **86**: 85–91.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314–331.

Boyd, J. M. (1953). The sheep population of Hirta, St. Kilda, 1952. *The Scottish Naturalist* **65**: 25–28.

Boyd, J. M. (1974). Introduction. In Jewell, P. A., Milner, C. and Boyd, J. M., eds., *Island Survivors: The Ecology of the Soay Sheep of St. Kilda*, pages 1–7, Athlone Press, London.

Boyd, J. M. and Boyd, I. L. (1990). *The Hebrides*. Collins, London.

Boyd, J. M., Doney, J. M., Gunn, R. G. and Jewell, P. A. (1964). The Soay sheep of the island of Hirta, St. Kilda: A study of a feral population. *Proceedings of the Zoological Society of London* **142**: 129–163.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**: 324–345.

Broad, T. E., Glass, B. C., Greer, G. J., Robertson, T. M., Bain, W. E., Lord, E. A. and McEwan, J. C. (2000). Search for a locus near to myostatin that increases muscling in Texel sheep in New Zealand. *Proceedings of the New Zealand Society of Animal Production* **60**: 110–112.

Brown, G. M., Leversha, M., Hulten, M., Ferguson-Smith, M. A., Affara, N. A. and Furlong, R. A. (1998). Genetic analysis of meiotic recombination in humans by use of sperm typing: Reduced recombination within a heterozygous paracentric inversion of chromosome 9q32-q34.3. *American Journal of Human Genetics* **62**: 1484–1492.

Burkin, D. J., Yang, F., Broad, T. E., Wienberg, J., Hill, D. F. and Ferguson-Smith, M. A. (1997). Use of the Indian muntjac idiogram to align conserved chromosomal segments in sheep and human genomes by chromosome painting. *Genomics* **46**: 143–147.

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**: 657–679.

Campbell, R. N. (1974). St. Kilda and its sheep. In Jewell, P. A., Milner, C. and Boyd, J. M., eds., *Island Survivors: The Ecology of the Soay Sheep of St. Kilda*, pages 8–35, Athlone Press, London.

Carlson, C. S., Eberle, M. A., Kruglyak, L. and Nickerson, D. A. (2004). Mapping complex disease loci in whole-genome association studies. *Nature* **429**: 446–452.

Casas, E., Shackelford, S. D., Keele, J. W., Stone, R. T., Kappes, S. M. and Koohmaraie, M. (2000). Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *Journal of Animal Science* **78**: 560–569.

Cheverud, J. M. (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87**: 52–58.

Christensen, K., Agerholm, J. S. and Larsen, B. (1992). Dairy breed bull with complex chromosome translocation: Fertility and linkage studies. *Hereditas* **117**: 199–202.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Clutton-Brock, T. and Pemberton, J., eds. (2004). *Soay Sheep: Dynamics and Selection in an Island Population.* Cambridge Univeristy Press, Cambridge.

Clutton-Brock, T. H., Grenfell, B. T., Coulson, T., MacColl, A. D. C., Illius, A., Forchhammer, M. C., Wilson, K., Lindström, J., Crawley, M. J. and Albon, S. D. (2004a). Population dynamics in Soay sheep. In Clutton-Brock, T. and Pemberton, J., eds., *Soay Sheep: Dynamics and Selection in an Island Population*, pages 52–88, Cambridge University Press, Cambridge.

Clutton-Brock, T. H., Illius, A., Wilson, K., Grenfell, B. T., MacColl, A. and Albon, S. D. (1997). Stability and instability in ungulate populations: An empirical analysis. *American Naturalist* **149**: 195–219.

Clutton-Brock, T. H., Pemberton, J. M., Coulson, T., Stevenson, I. R. and MacColl, A. D. C. (2004b). The sheep of St Kilda. In Clutton-Brock, T. and Pemberton, J., eds., *Soay Sheep: Dynamics and Selection in an Island Population*, pages 17–51, Cambridge University Press, Cambridge.

Clutton-Brock, T. H., Price, O. F., Albon, S. D. and Jewell, P. A. (1991). Persistant instability and population regulation in the Soay sheep. *Journal of Animal Ecology* **60**: 593–608.

Clutton-Brock, T. H., Price, O. F., Albon, S. D. and Jewell, P. A. (1992). Early development and population fluctuations in Soay sheep. *Journal of Animal Ecology* **61**: 381–396.

Clutton-Brock, T. H., Stevenson, I. R., Marrow, P., MacColl, A. D., Houston, A. I. and McNamara, J. M. (1996). Population fluctuations, reproductive costs and life history tactics in female Soay sheep. *Journal of Animal Ecology* **65**: 675–689.

Cobb, J. P., Mindrinos, M. N., Miller-Graziano, C., Calvano, S. E., Baker, H. V., Xiao, W., Laudanski, K., Brownstein, B. H., Elson, C. M., Hayden, D. L., Herndon, D. N., Lowry, S. F., Maier, R. V., Schoenfeld, D. A., Moldawer, L. L., Davis, R. W., Tompkins, R. G. and Inflammation

and Host Response to Injury Large-Scale Collaborative Research Program (2005). Application of genome-wide expression analysis to human health and disease. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4801–4806.

Cockett, N. E., Jackson, S. P., Shay, T. L., Farnir, F., Berghmans, S., Snowder, G. D., Nielsen, D. M. and Georges, M. (1996). Polar overdominance at the ovine callipyge locus. *Science* **273**: 236–238.

Cockett, N. E., Jackson, S. P., Shay, T. L., Nielson, D., Moore, S. S., Steele, M. R., Barendse, W., Green, R. D. and Georges, M. (1994). Chromosomal localization of the callipyge gene in sheep (Ovis aries) using bovine DNA markers. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 3019–3023.

Cockett, N. E., Smit, M. A., Bidwell, C. A., Segers, K., Hadfield, T. L., Snowder, G. D., Georges, M. and Charlier, C. (2005). The *callipyge* mutation and other genes that affect muscle hypertrophy in sheep. *Genetics Selection Evolution* **37 (Suppl. 1)**: S65–S81.

Coltman, D. W. and Pemberton, J. M. (2004). Inheritance of coat colour and horn type in Hirta Soay sheep. In Clutton-Brock, T. and Pemberton, J., eds., *Soay Sheep: Dynamics and Selection in an Island Population*, pages 321–327, Cambridge University Press, Cambridge.

Coltman, D. W., Pilkington, J. G. and Pemberton, J. M. (2003). Fine-scale genetic structure in a free-living ungulate population. *Molecular Ecology* **12**: 733–742.

Coppieters, W., Riquet, J., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. and Georges, M. (1998). A QTL with major effect on milk yeild and composition maps to bovine chromosome 14. *Mammalian Genome* **9**: 540–544.

Coulson, T., Albon, S., Pilkington, J. and Clutton-Brock, T. (1999). Small-scale spatial dynamics in a fluctuating ungulate population. *Journal of Animal Ecology* **68**: 658–671.

Coulson, T., Catchpole, E. A., Albon, S. D., Morgan, B. J. T., Pemberton, J. M., Clutton-Brock, T. H., Crawley, M. J. and Grenfell, B. T. (2001). Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* **292**: 1528–1531.

Crawford, A. M. and Cuthbertson, R. P. (1996). Mutations in sheep microsatellites. *Genome Research* **6**: 876–879.

Crawford, A. M., Dodds, K. G., Ede, A. J., Pierson, C. A., Montgomery, G. W., Garmonsway, H. G., Beattie, A. E., Davies, K., Maddox, J. F., Kappes, S. W., Stone, R. T., Nguyen, T. C., Penty, J. M., Lord, E. A., Broom,

J. E., Buitkamp, J., Schwaiger, W., Epplen, J. T., Matthew, P., Matthews, M. E., Hulme, D. J., Beh, K. J., McGraw, R. A. and Beattie, C. W. (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* **140**: 703–724.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**: 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**: 33–43.

Davis, G. H. (2005). Major genes affecting ovulation rate in sheep. *Genetics Selection Evolution* **37 (Suppl. 1)**: S11–S23.

Davis, G. H., Dodds, K. G., Wheeler, R. and Jay, N. P. (2001). Evidence that an imprinted gene on the X chromosome increases ovulation rate in sheep. *Biology of Reproduction* **64**: 216–221.

Davis, G. H., McEwan, J. C., Fennessy, P. F. and Dodds, K. G. (1995). Discovery of the Inverdale gene (*FecX*). *Proceedings of the New Zealand Society of Animal Production* **55**: 289–290.

Davis, G. H., McEwan, J. C., Fennessy, P. F., Dodds, K. G. and Farquhar, P. A. (1991). Evidence for the presence of a major gene influencing ovulation rate on the X chromosome of sheep. *Biology of Reproduction* **44**: 620–624.

Davis, G. H., Montgomery, G. W., Allison, A. J., Kelly, R. W. and Bay, A. R. (1982). Segregation of a major gene influencing fecundity in progeny of Booroola sheep. *New Zealand Journal of Agricultural Research* **25**: 525–529.

de Gooijer, J. G. (2001). Cross-validation criteria for SETAR model selection. *Journal of Time Series Analysis* **22**: 267–281.

de Gortari, M. J., Freking, B. A., Cuthbertson, R. P., Kappes, S. M., Keele, J. W., Stone, R. T., Leymaster, K. A., Dodds, K. G., Crawford, A. M. and Beattie, C. W. (1998). A second-generation linkage map of the sheep genome. *Mammalian Genome* **9**: 204–209.

de Koning, D. J., Pong-Wong, R., Varona, L., Evans, G. J., Giuffra, E., Sanchez, A., Plastow, G., Noguera, J. L., Andersson, L. and Haley, C. S. (2003). Full pedigree quantitative trait locus analysis in commercial pigs using variance components. *Journal of Animal Science* **81**: 2155–2163.

Dekkers, J. C. M. and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**: 22–32.

Dentine, M. R. and Cowan, C. M. (1990). An analytical model for the estimation of chromosome substitution effects in the offspring of individuals

heterozygous at a segregating marker locus. *Theoretical and Applied Genetic* **79**: 775–780.

Devlin, B. and Risch, N. (1995). A comparison of linkage disequilibrium measures for fine mapping. *Genomics* **29**: 311–322.

Dodds, K. G., Ball, R., Djorovic, N. and Carson, S. D. (2004). The effect of an imprecise map on interval mapping QTLs. *Genetical Research* **84**: 47–55.

Dodds, K. G., Tate, M. L., McEwan, J. C. and Crawford, A. M. (1996). Exclusion probabilities for pedigree testing farm animals. *Theoretical and Applied Genetics* **92**: 966–975.

Doerge, R. W. and Rebaï, A. (1996). Significance thresholds for QTL interval mapping tests. *Heredity* **76**: 459–464.

Dolling, C. H. S. (1961). Hornedness and polledness in sheep. IV. Triple alleles affecting horn growth in the Merino. *Australian Journal of Agricultural Research* **12**: 356–361.

Doney, J. M., Ryder, M. L., Gunn, R. G. and Grubb, P. (1974). Colour, conformation, affinities, fleece and patterns of inheritance of the Soay sheep. In Jewell, P. A., Milner, C. and Boyd, J. M., eds., *Island Survivors: The Ecology of the Soay Sheep of St. Kilda*, pages 88–125, Athlone Press, London.

Douch, P. G. C., Green, R. S., Morris, C. A., Bisset, S. A., Vlassoff, A., Baker, R. L., Watson, T. G., Hurford, A. P. and Wheeler, M. (1995a). Genetic and phenotypic relationships among anti-*Trichostrongylus colubriformis* antibody level, faecal egg count and body weight traits in grazing Romney sheep. *Livestock Production Science* **41**: 121–132.

Douch, P. G. C., Green, R. S., Morris, C. A. and Hickey, S. M. (1995b). Genetic factors affecting antibody responses to four species of nematode parasite in Romney ewe lambs. *International Journal for Parasitology* **25**: 823–828.

Ducos, A., Pinton, A., Yerle, M., Séguéla, A., Berland, H.-M., Brun-Baronnat, C., Bonnet, N. and Darré, R. (2002). Cytogenetic and molecular characterization of eight new reciprocal translocations in the pig species. Estimation of their incidence in French populations. *Genetics Selection Evolution* **34**: 389–406.

Dunning, A. M., Durocher, F., Healey, C. S., Teare, M. D., McBride, S. E., Carlomagno, F., Xu, C.-F., Dawson, E., Rhodes, S., Ueda, S., Lai, E., Luben, R. N., Van Rensberg, E. J., Mannermaa, A., Kataja, V., Rennart, G., Dunham, I., Purvis, I., Easton, D. and Ponder, B. A. J. (2000). The extent of linkage disequilibrium in four populations with distinct demographic histories. *American Journal of Human Genetics* **67**: 1544–1554.

Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**: 373–386.

Eaves, I. A., Merriman, T. R., Barber, R. A., Nutland, S., Tuomilehto-Wolf, E., Tuomilehto, J., Cucca, F. and Todd, J. A. (2000). The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nature Genetics* **25**: 320–322.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**: 1–26.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68**: 589–599.

Elo, K. T., Vilkki, J., de Koning, D. J., Velmala, R. J. and Mäki-Tanila, A. V. (1999). A quantitative trait locus for live weight maps to bovine chromosome 23. *Mammalian Genome* **10**: 831–835.

Emigh, T. H. and Pollack, E. (1979). Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theoretical Population Biology* **15**: 86–107.

Everts-van der Wind, A., Kata, S. R., Band, M. R., Rebeiz, M., Larkin, D. M., Everts, R. E., Green, C. A., Liu, L., Natarajan, S., Goldammer, T., Lee, J. H., McKay, S., Womack, J. E. and Lewin, H. A. (2004). A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. *Genome Research* **14**: 1424–1437.

Eyre-Walker, A. (2000). Do mitochondria recombine in humans? *Philosophical Transactions of the Royal Society of London. Series B* **355**: 1573–1580.

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. and Georges, M. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**: 220–227.

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilkki, J. and Georges, M. (2002). Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**: 275–287.

Felsenstein, J. (1971). Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**: 581–597.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Claredon Press, Oxford.

Forchhammer, M. C., Clutton-Brock, T. H., Lindström, J. and Albon, S. D. (2001). Climate and population density induce long-term cohort variation in a northern ungulate. *Journal of Animal Ecology* **70**: 721–729.

Frankham, R. (1995a). Effective population size/adult population size ratios in wildlife: A review. *Genetical Research* **66**: 95–107.

Frankham, R. (1995b). Inbreeding and extinction: A threshold effect. *Conservation Biology* **9**: 792–799.

Franklin, I. R. (1997). Systematics and phylogeny of the sheep. In Piper, L. and Ruvinsky, A., eds., *The Genetics of Sheep*, pages 1–12, CAB International, Cambridge.

Freking, B. A., Keele, J. W., Beattie, C. W., Kappes, S. M., Smith, T. P. L., Sonstegard, T. S., Nielsen, M. K. and Leymaster, K. A. (1998). Evaluation of the ovine *callipyge* locus. I. Relative chromosomal position and gene action. *Journal of Animal Science* **76**: 2062–2071.

Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L. and Smith, T. P. L. (2002). Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Research* **12**: 1496–1506.

Galloway, S. M., Hanrahan, V., Dodds, K. G., Potts, M. D., Crawford, A. M. and Hill, D. F. (1996). A linkage map of the ovine X chromosome. *Genome Research* **6**: 667–677.

Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P. E., Juengel, J. L., Jokiranta, T. S., McLaren, R. J., Luiro, K., Dodds, K. G., Montgomery, G. W., Beattie, A. E., Davis, G. H. and Ritvos, O. (2000). Mutations in an oocyte-derived growth factor gene (*BMP15*) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics* **25**: 279–283.

George, A. W., Visscher, P. M. and Haley, C. S. (2000). Mapping quantitative trait loci in complex pedigrees: A two-step variance component approach. *Genetics* **156**: 2081–2092.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. and Thompson, R. (2002). *ASReml User Guide Release 1.0.* VSN International Ltd, Hemel Hempstead.

Goddard, M. E. (1991). Mapping genes for quantitative traits using linkage disequilibrium. *Genetics Selection Evolution* **23 (Suppl. 1)**: S131–S134.

Goldgar, D. E. (1990). Multipoint analysis of human quantitative genetic variation. *American Journal of Human Genetics* **47**: 957–967.

Göring, H. H. H., Terwilliger, J. D. and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *American Journal of Human Genetics* **69**: 1357–1369.

Grapes, L., Dekkers, J. C. M., Rothschild, M. F. and Fernando, R. L. (2004). Comparing linkage disequilibrium methods for fine mapping quantitative trait loci. *Genetics* **166**: 1561–1570.

Green, P., Falls, K. and Crooks, S. (1990). *Cri-Map Version 2.4*. School of Medicine, Washington University, St. Louis, Mo. USA.

Grenfell, B. T., Price, O. F., Albon, S. D. and Clutton-Brock, T. H. (1992). Overcompensation and population cycles in an ungulate. *Nature* **355**: 823–826.

Grenfell, B. T., Wilson, K., Finkenstadt, B. F., Coulson, T. N., Murray, S., Albon, S. D., Pemberton, J. M., Clutton-Brock, T. H. and Crawley, M. J. (1998). Noise and determinism in synchronized sheep dynamics. *Nature* **394**: 674–677.

Grignola, F. E., Hoeschele, I., Zhang, Q. and Thaller, G. (1996). Mapping quantitative trait loci in outcross populations via residual maximum likelihood. II. A simulation study. *Genetics Selection Evolution* **28**: 491–504.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. and Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Research* **12**: 222–231.

Grobet, L., Poncelet, D., Royo, L. J., Brouwers, D., Pirottin, D., Michaux, C., Ménissier, F., Zanotti, M., Dunner, S. and Georges, M. (1998). Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mammalian Genome* **9**: 210–213.

Groenen, M. A. M., Cheng, H. H., Bumstead, N., Benkel, B. F., Briles, W. E., Burke, T., Burt, D. W., Crittenden, L. B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A. P., Soller, M., Takahashi, H. and Vignal, A. (2000). A concensus linkage map of the chicken genome. *Genome Research* **10**: 137–147.

Grubb, P. (1974). Population dynamics of the Soay sheep. In Jewell, P. A., Milner, C. and Boyd, J. M., eds., *Island Survivors: The Ecology of the Soay Sheep of St. Kilda*, pages 242–272, Athlone Press, London.

Gulland, F. M. D. (1992). Role of nematode infections in mortality of Soay sheep (*Ovis aries* L.) during a population crash on St. Kilda. *Parasitology* **105**: 493–503.

Gustavsson, I. (1980). Chromosome abberations and their influence on the reproductive performance of domestic animals - A review. *Zeitschrift für Tierzüchtung und Züchtungsbiologie* **97**: 176–195.

Guttenbach, M., Engel, W. and Schmid, M. (1997). Analysis of structural and numerical chromosome abnormalities in sperm of normal men and carriers of constitutional chromosome aberrations: A review. *Human Genetics* **100**: 1–21.

Haley, C. (1999). Advances in quantitative trait locus mapping. In Dekkers, J. C. M., Lamont, S. J. and Rothschild, M. F., eds., *From J. L. Lush To Genomics: Visions for Animal Breeding and Genetics*, pages 47–59, Iowa State University, Ames.

Haley, C. S. and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.

Haley, C. S., Knott, S. A. and Elsen, J.-M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**: 1195–1207.

Hanrahan, J. P. (1991). Evidence for single gene effects on ovulation rate in the Cambridge and Belcare breeds. In Elsen, J. M., Bodin, L. and Thimonier, J., eds., *Major Genes for Reproduction in Sheep*, pages 93–102, INRA, Paris.

Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R. and Galloway, S. M. (2004). Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belcare sheep (*Ovis aries*). *Biology of Reproduction* **70**: 900–909.

Harris, R. B. and Allendorf, F. W. (1989). Genetically effective population size of large mammals: An assessment of estimators. *Conservation Biology* **3**: 181–191.

Hayes, B. and Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitatie traits in livestock. *Genetics Selection Evolution* **33**: 209–229.

Hayes, B. J., Visscher, P. M., McPartlan, H. C. and Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* **13**: 635–643.

Heath, S. C. (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics* **61**: 748–760.

Hedrick, P. (1987). Gametic disequilibrium measures: Proceed with caution. *Genetics* **117**: 331–341.

Henricson, B. and Bäckström, L. (1964). Translocation heterozygosity in a boar. *Hereditas* **52**: 166–170.

Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., VanRaden, P. M. and Lewin, H. A. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological genomics* **1**: 165–175.

Hiendleder, S., Kaupe, R., B.and Wassmuth and Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London B* **269**: 893–904.

Hiendleder, S., Mainz, K., Plante, Y. and Lewalski, H. (1998). Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: No evidence for contributions from urial and argali sheep. *Journal of Heredity* **89**: 113–120.

Hill, W. G. (1972a). Effective size of populations with overlapping generations. *Theoretical Population Biology* **3**: 278–289.

Hill, W. G. (1972b). Estimation of genetic change. I. General theory and design of control populations. *Animal Breeding Abstracts* **40**: 1–15.

Hill, W. G. (1979). A note on effective population size with overlapping generations. *Genetics* **92**: 317–322.

Hill, W. G. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**: 226–231.

Hill, W. G. and Weir, B. S. (1994). Maximum-likelihood estimation of gene location by linkage disequilibrium. *American Journal of Human Genetics* **54**: 705–714.

Hospital, F. and Charcosset, A. (1997). Marker-assisted introgression of quantitative trait loci. *Genetics* **147**: 1469–1485.

Hu, N., Wang, C., Hu, Y., Yang, H. H., Giffen, C., Tang, Z.-Z., Han, X.-Y., Goldstein, A. M., Emmert-Buck, M. R., Buetow, K. H., Taylor, P. R. and Lee, M. P. (2005). Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Research* **65**: 2542–2546.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.

Iannuzzi, L., Di Meo, G. P., Perucatti, A. and Incarnato, D. (1999). Comparison of the human with the sheep genomes by use of human chromosome-specific painting probes. *Mammalian Genome* **10**: 719–723.

Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., Hirano, T., Itoh, T., Watanabe, T., Reed, K. M., Snelling, W. M., Kappes, S. M., Beattie, C. W., Bennett, G. L. and Sugimoto, Y. (2004). A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* **14**: 1987–1998.

Iturria, S. J., Williams, J. T., Almasy, L., Dyer, T. D. and Blangero, J. (1999). An empilical test of the significance of an observed quantitative trait locus effect the preserves additive genetic variation. *Genetic Epidemiology* **17 (Suppl. 1)**: S169–S173.

Jackson, S. P., Miller, M. F. and Green, R. D. (1997). Phenotypic characterization of Rambouillet sheep expression the *Callipyge gene*: III. Muscle weights and muscle weight distribution. *Journal of Animal Science* **75**: 133–138.

Janssen, M., weimann, C., Gauly, M. and Erhardt, G. (2002). Associations between infections with *Haemonchus contortus* and genetic markers on ovine chromsome 20. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpelleir, France.

Jeon, J.-T., Carlborg, Ö., Törnsten, A., Giuffra, E., Amager, V., Chardon, P., Andersson-Eklind, L., Andersson, K., Hansson, I., Lundström, K. and Andersson, L. (1999). A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nature Genetics* **21**: 157–158.

Jewell, P. A., Milner, C. and Boyd, J. M., eds. (1974). *Island Survivors: The Ecology of the Soay Sheep of St. Kilda.* Athlone Press, London.

Joerg, H., Garner, D., Rieder, S., Suwattana, D. and Stranzinger, G. (2001). Molecular genetic characterization of Robertsonian translocations in cattle. *Journal of Animal Breeding and Genetics* **118**: 371–377.

Johnson, D. L. (1977). Inbreeding in populations with overlapping generations. *Genetics* **87**: 581–591.

Jopson, N. B., Nicholl, G. B., Stevenson-Barry, J. M., Duncan, S., Greer, G. J., Bain, W. E., Gerard, E. M., Glass, B. C., Broad, T. E. and McEwan, J. C. (2001). Mode of inheritance and effects on meat quality of the *rib-eye muscling (REM)* QTL in sheep. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* **14**: 111–114.

Jorgenson, E., Tang, H., Gadde, M., Province, M., Leppert, M., Kardia, S., Schork, N., Cooper, R., Rao, D. C., Boerwinkle, E. and Risch, N. (2005). Ethnicity and human genetic linkage maps. *American Journal of Human Genetics* **76**: 276–290.

Jung, Y. C., Rothschild, M. F., Flanagan, M. P., Christian, L. L. and Warner, C. M. (1989). Association of restriction fragment length polymorphisms of swine leucocyte class I genes with production traits of Duroc and Hampshire boars. *Animal Genetics* **20**: 79–91.

Kalousek, D. K. and Lau, A. E. (1992). Pathology of spontaneous abortion. In Dimmick, J. E. and Kalousek, D. K., eds., *Developmental Pathology of the Embryo and Fetus*, pages 55–82, Lippincot, Philadelphia.

Kappes, S. M., Keele, J. W., Stone, R. T., McGraw, R. A., Sonstegard, T. S., Smith, T. P., Lopez-Corrales, N. L. and Beattie, C. W. (1997). A second-generation linkage map of the bovine genome. *Genome Research* **7**: 235–249.

Kim, K. S., Larsen, N., Short, T., Plastow, G. and Rothschild, M. F. (2000). A missence variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth and feed intake traits. *Mammalian Genome* **11**: 131–135.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

Kimura, M. and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.

Kmiec, M. (1999a). Transferrin polymorphism versus growth fate in lambs, Polish long-wool sheep. I. Frequency of genes and genotypes of transferrin in flocks of Polish long-wool sheep. *Archiv Für Tierzucht - Archives of Animal Breeding* **42**: 393–402.

Kmiec, M. (1999b). Transferrin polymorphism versus growth rate in lambs, Polish long-wool sheep. II. Analysis of relation between transferrin polymorphism of lamb blood serum versus growth rate up to age of 5 months. *Archiv Für Tierzucht - Archives of Animal Breeding* **42**: 469–479.

Knott, S. A., Elsen, J. M. and Haley, C. S. (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**: 71–80.

Kong, A., Barnard, J., Gudbjartsson, D. F., Thorleifsson, G., Jonsdottir, G., Sigurdardottir, S., Richardsson, B., Jonsdottir, J., Thorgeirsson, T., Frigge, M. L., Lamb, N. E., Sherman, S., Gulcher, J. R. and Stefansson, K. (2004). Recombination rate and reproductive success in humans. *Nature Genetics* **36**: 1203–1206.

Koohmaraie, M., Shackelford, S. D., Wheeler, T. L., Lonergan, S. M. and Doumit, M. E. (1995). A muscle hypertrophy condition in lamb (Callipgye): Characterization of effects on muscle growth and meat quality traits. *Journal of Animal Science* **73**: 3596–3607.

Kruglyak, L. (1999). Prospects for whole genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139–144.

Kruglyak, L. and Lander, E. S. (1995). A nonparametric approach for mapping quantitative trait loci. *Genetics* **139**: 1421–1428.

Lande, R. and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**: 743–756.

Lander, E. and Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**: 241–247.

Lander, E. S. (1999). Array of hope. *Nature Genetics* **21 (Suppl.)**: 3–4.

Lander, E. S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.

Lange, K., Cantor, R., Horvath, S., Perola, M., Sabatti, C., Sinsheimer, J. and Sobel, E. (2001). Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *American Journal of Human Genetics* **69 (Suppl.)**: A1866.

Larkin, D. M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P. A., Bachman, S., Green, C., Wright, C. L., Campos, E. J., Benson, L. D., Edwards, J., Liu, L., Osoegawa, K., Womack, J. E., de Jong, P. J. and Lewin, H. A. (2003). A cattle-human comparative map built with cattle BAC-ends and human genome sequence. *Genome Research* **13**: 1966–1972.

Lecerf, F., Mulsant, P., Elsen, J. M. and Bodin, L. (2002). Localization and mapping of a major gene controlling ovulation rate in Lacune sheep. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France.

Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations; Heterotic models. *Genetics* **49**: 49–67.

Lewontin, R. C. (1988). On measures of gametic disequilibrium. *Genetics* **120**: 849–852.

Lindström, J., Coulson, T., Kruuk, L., Forchhammer, M. C., Coltman, D. W. and Clutton-Brock, T. H. (2002). Sex-ratio variation in Soay sheep. *Behavioral Ecology and Sociobiology* **53**: 25–30.

Loftus, R. and Scherf, B., eds. (1993). *World Watch List for Domestic Animal Diversity*. FAO, Rome.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland.

Ma, R. Z., Beever, J. E., Da, Y., Green, C. A., Russ, I., Park, C., Heyen, D. W., Everts, R. E., Fisher, S. R., Overton, K. M., Teale, A. J., Kemp, S. J., Hines, H. C., Guéring, G. and Lewin, H. A. (1996). A male linkage map of the cattle (*Bos taurus*) genome. *Journal of Heredity* **87**: 261–271.

Machado, M. B. B., Alencar, M. M., Pereira, A. P., Oliveira, H. N., Casas, E., Coutinho, L. L. and Regitano, L. C. A. (2003). QTL affecting body weight in a candidate region of cattle chromosome 5. *Genetics and Molecular Biology* **26**: 259–265.

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Reviews Genetics* **35**: 303–339.

Maddox, J. (2003). Australian Sheep Gene Mapping Web Site - Current sex averaged, female, male distances for best position maps (v4.2). <http://rubens.its.unimelb.edu.au/~jillm/jill.htm>.

Maddox, J. F. (2004). A presentation of the differences between the sheep and goat genetic maps. *Genetics Selection Evolution* **37 (Suppl. 1)**: S1–S10.

Maddox, J. F., Davies, K. P., Crawford, A. M., Hulme, D. J., Vaiman, D., Cribiu, E. P., Freking, B. A., Beh, K. J., Cockett, N. E., Kang, N., Riffkin, C. D., Drinkwater, R., Moore, S. S., Dodds, K. G., Lumsden, J. M., van Stijn, T. C., Phua, S. H., Adelson, D. L., Burkin, H. R., Broom, J. E., Buitkamp, J., Cambridge, L., Cushwa, W. T., Gerard, E., Galloway, S. M., Harrison, B., Hawken, R. J., Hiendleder, S., Henry, H. M., Medrano, J. F., Paterson, K. A., Schibler, L., Stone, R. T. and van Hest, B. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research* **11**: 1275–1289.

Mangin, B., Goffinet, B. and Rebaï, A. (1994). Constructing confidence intervals for QTL location. *Genetics* **138**: 1301–1308.

Maraganore, D. M., de Andrade, M., Lesnick, T. G., Strain, K. J., Farrer, M. J., Rocca, W. A., Pant, P. V. K., Frazer, K. A., Cox, D. R. and Ballinger, D. G. (2005). High-resolution whole-genome association study of Parkinson disease. *American Journal of Human Genetics* **77**: 685–693.

Marcq, F., Elsen, J.-M., El Barkouki, S., Bouix, J., Eychenne, F., Grobet, L., Karim, L., Laville, E., Nezer, C., Royo, L., Sayd, T., Bibe, B., Leroy, P. L. and Georges, M. (1998). Investigating the role of myostatin in the determinism of double muscling characterizing Belgian Texel sheep. *Animal Genetics* **29 (Suppl. 1)**: 52–53.

Marcq, F., Elsen, J. M., Marot, V., Bouix, J., Coppieters, W., Eychenne, F., Laville, E., Nezer, C., Sayd, T., Bibè, B., Georges, M. and Leroy, P. L. (1999). Mapping quantitative trait loci causing the muscular hypertrophy of Belgian Texel sheep. In *Proceedings of 50th Annual Meeting of the European Association for Animal Production*, Zürich, Switzerland.

Martin, R. H., Chernos, J. E., Lowry, R. B., Pattinson, H. A., Barclay, L. and Ko, E. (1994). Analysis of sperm chromosome complements from a man heterozygous for a paracentric inversion of chromosome 1. *Human Genetics* **93**: 135–138.

Mason, I. L. (1996). *A World Dictionary of Livestock Breeds, Types and Varieties.* CAB International, Wallingford, 4th edition.

McClintock, B. (1931). Cytological observations of deficiencies involving known genes, translocations and an inversion in *Zea mays. Missouri Agricultural Experiment Station Research Bulletin* **163**: 1–30.

McClintock, B. (1933). The association of non-homologous parts of chromosomes in the mid-prophase of meiosis in *Zea mays. Zeitschrift fur Zellforschung und Mikroskopische Anatomie* **19**: 191–237.

McEwan, J. C., Dodds, K. G., Greer, G. J., Bain, W. E., Duncan, S. J., Wheeler, R., Knowler, K. J., Reid, P. J., Green, R. S. and Douch, P. G. C. (1994). Genetic estimates for parasite resistance traits in sheep and their correlations with production traits. In *Proceedings of the 23rd Annual Meeting of the New Zealand Society for Parasitology*, page 4, Wallaceville.

McRae, A. F., McEwan, J. C., Dodds, K. G., Wilson, T., Crawford, A. M. and Slate, J. (2002). Linkage disequilibrium in domestic sheep. *Genetics* **160**: 1113–1122.

Meuwissen, T. H. E. and Goddard, M. E. (2000). Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* **155**: 421–430.

Meuwissen, T. H. E. and Goddard, M. E. (2001). Prediction of identity by descent probabilities from marker haplotypes. *Genetics Selection Evolution* **33**: 605–634.

Meuwissen, T. H. E., Karlsen, A., Lien, S., Olsaker, I. and Goddard, M. E. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**: 373–379.

Milner, J. M., Elston, D. and Albon, S. D. (1999). Estimating the contributions of population density and climatic factors to interannual variation in survival of Soay sheep. *Journal of Animal Ecology* **68**: 1235–1247.

Montgomery, G. W., Crawford, A. M., Penty, J. M., Dodds, K. G., Ede, A. J., Henry, H. M., Pierson, C. A., Lord, E. A., Galloway, S. M., Schmack, A. E., Sise, J. A., Swarbrick, P. A., Hanrahan, V., Buchanan, F. C. and Hill, D. F. (1993). The ovine Booroola fecundity gene (*FecB*) is linked to markers from a region of human chromosome 4q. *Nature Genetics* **4**: 410–414.

Montgomery, G. W., Henry, H. M., Dodds, K. G., Beattie, A. E., Wuliji, T. and Crawford, A. M. (1996). Mapping of the *Horns* (*Ho*) locus in sheep: A

further locus controlling horn development in domestic animals. *Journal of Heredity* **87**: 358–363.

Montgomery, G. W., Lord, E. A., Penty, J. M., Dodds, K. G., Broad, T. E., Cambridge, L., Sunden, S. L. F., Stone, R. T. and Crawford, A. M. (1994). The Booroola fecundity (*FecB*) gene maps to sheep chromosome 6. *Genomics* **22**: 148–153.

Moody, D. E., Pomp, D., Newman, S. and MacNeil, M. D. (1996). Characterization of DNA polymorphism in three populations of Hereford cattle and their associations with growth and maternal EPD in line 1 Herefords. *Journal of Animal Science* **74**: 1784–1793.

Moran, P. A. P. (1962). *The Statistical Processes of Evolutionary Biology.* Clarendon Press, Oxford.

Morris, C. A., Towers, N. R., Hohenboken, W. D., Maqbool, N., Smith, B. L. and Phua, S. H. (2004). Inheritance of resistance to facial eczema: A review of research findings from sheep and cattle in New Zealand. *New Zealand Veterinary Journal* **52**: 205–215.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* **7**: 277–318.

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet, J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognié, Y., Chitour, N. and Elsen, J.-M. (2001). Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 5104–5109.

Nachman, M. W. and Crowell, S. L. (2000). Estimation of the mutation rate per nucleotide in humans. *Genetics* **156**: 297–304.

Nagamine, Y., Haley, C. S., Sewalem, A. and Visscher, P. M. (2003). Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics* **164**: 629–635.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**: 945–966.

Nettleton, D. and Doerge, R. W. (2000). Accounting for variability in the use of permutation testing to detect quantitative trait loci. *Biometrics* **56**: 52–58.

Nezer, C., Collette, C., Moreau, L., Brouwers, B., Kim, J.-J., Giuffra, E., Buys, N., Andersson, L. and Georges, M. (2003). Haplotype sharing refines the location of an imprinted quantitative trait locus with major effect on muscle mass to a 250-kb chromosome segment containing the porcine *IGF2* gene. *Genetics* **165**: 277–285.

Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., Karim, L., Kvasz, A., Leroy, P. and Georges, M. (1999). An imprinted QTL with major effect in muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nature Genetics* **21**: 155–156.

Nicoll, G. B., Burkin, H. R., Broad, T. E., Jopson, N. B., Greer, G. J., Bain, W. E., Wright, C. S., Dodds, K. G., Fennessy, P. F. and McEwan, J. C. (1998). Genetic linkage of microsatellite markers to the Carwell locus for rib-eye muscling in sheep. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production* **26**: 529–532.

Nsengimana, J., Baret, P., Haley, C. S. and Visscher, P. M. (2004). Linkage disequilibrium in the domesticated pig. *Genetics* **166**: 1395–1404.

Nunney, L. (1991). The influence of age structure and fecundity on effective population size. *Proceedings of the Royal Society of London B* **246**: 71–76.

Nunney, L. (1993). The influence of mating system and overlapping generations in effective population size. *Evolution* **47**: 1329–1341.

O'Connell, J. R. and Weeks, D. E. (1998). PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *American Journal of Human Genetics* **63**: 259–266.

Oliver, F., Christians, J. K., Liu, X., Rhind, S., Verma, V., Davison, C., Brown, S. D. M., Denny, P. and Keightly, P. D. (2005). Regulatory variation at Glypican-3 underlies a major growth QTL in mice. *PLoS Biology* **3**: 872–877.

Park, C., Frank, M. T. and Lewin, H. A. (1999). Fine-mapping of a region of variation in recombination rate on BTA23 to the *D23S22-D23S23* interval using sperm typing and meiotic breakpoint analysis. *Genomics* **59**: 143–149.

Park, C., Russ, I., Da, Y. and Lewin, H. A. (1995). Genetic mapping of *F13A* to BTA23 by sperm typing: Difference in recombination rate between bulls in the *DYA-PRL* interval. *Genomics* **27**: 113–118.

Paterson, K. A., McEwan, J. C., Dodds, K. G., Morris, C. A. and Crawford, A. M. (2001). Fine mapping a locus affecting host resistance to internal parasites in sheep. *Proceedings of the Association for Advancement in Animal Breeding and Genetics* **14**: 91–94.

Paterson, S., Wilson, K. and Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences of the United States of America* **95**: 3714–3719.

Pemberton, J. M., Coltman, D. W., Smith, J. A. and Bancroft, D. R. (2004). Mating patterns and male breeding success. In Clutton-Brock, T. and Pemberton, J., eds., *Soay Sheep: Dynamics and Selection in an Island Population*, pages 166–189, Cambridge University Press, Cambridge.

Pemberton, J. M., Coltman, D. W., Smith, J. A. and Pilkinton, J. G. (1999). Molecular analysis of a promiscuous, fluctuating mating system. *Biological Journal of the Linnean Society* **68**: 289–301.

Phua, S. H., Dodds, K. G., Morris, C. A., Paterson, K. A., McEwan, J. C., Garmonsway, H. G., Towers, N. R. and Crawford, A. M. (1999). Catalase gene is associated with facial eczema disease resistance in sheep. *Animal Genetics* **30**: 286–295.

Piepho, H.-P. (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* **157**: 425–432.

Pinton, A., Ducos, A., Berland, H. M., Séguéla, A., Blanc, M. F., Darré, A., Mimar, S. and Darré, R. (1997). A new Robertsonian translocation in the Holstein breed. *Genetics Selection Evolution* **29**: 523–526.

Piper, L. R. and Bindon, B. M. (1982). The Booroola Merino and the performance of medium non-Peppin crosses at Armidale. In Piper, L. R., Bindon, B. M. and Nethery, R. D., eds., *The Booroola Merino*, pages 9–20, CSIRO, Melbourne.

Piper, L. R., Bindon, B. M. and Davis, G. H. (1985). The single gene inheritance of the high litter size of the Booroola Merino. In Piper, L. R., Bindon, B. M. and Nethery, R. D., eds., *Genetics of Reproduction in Sheep*, pages 115–125, Butterworths, London.

Popescu, C. P. and Boscher, J. (1986). A new reciprocal translocation in a hypoprolific boar. *Genetics Selection Evolution* **18**: 123–130.

Pritchard, J. K. and Przeworski, M. (2001). Linkage disequilibrium in humans: Models and data. *American Journal of Human Genetics* **69**: 1–14.

Purvis, I. W. and Franklin, I. R. (2005). Major genes and QTL influencing wool production and quality: A review. *Genetics Selection Evolution* **37 (Suppl. 1)**: S97–S107.

Ramos, P. S., Bitgood, J. J. and Ponce de Leon, A. (1999). Novel chromosomal translocation in chicken uncovered by double colour FISH. *Animal Biotechnology* **10**: 119–122.

Rebaï, A., Goffinet, B. and Mangin, B. (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**: 235–240.

Rebaï, A., Goffinet, B. and Mangin, B. (1995). Comparing power of different methods for QTL detection. *Biometrics* **51**: 87–99.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature* **411**: 199–204.

Rohrer, G. A., Alexander, L. J., Hu, Z., Smith, T. P. L., Keele, J. W. and Beattie, C. W. (1996). A comprehensive map of the porcine genome. *Genome Research* **6**: 371–391.

Ryder, M. L. (1984). Sheep. In Mason, I. L., ed., *Evolution of Domesticated Animals*, pages 63–85, Longman, London.

Schibler, L., Vaiman, D., Oustry, A., Giraud-Delville, C. and Cribiu, E. P. (1998). Comparative gene mapping: A fine scale survey of chromosome rearrangements between ruminants and humans. *Genome Research* **8**: 901–915.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J. M., Wimmers, K., Burt, D. W., Smith, J., A'Hara, S., Law, A., Griffin, D. K., Bumstead, N., Kaufman, J., Thomson, P. A., Burke, T., Groenen, M. A. M., Crooijmans, R. P. M. A., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Tixier-Boichard, M., Ladjali-Mohammedi, K., Hillel, J., Mäki-Tanila, A., Cheng, H. H., Delany, M. E., Burnside, J. and Mizuno, S. (2000). First report on chicken genes and chromosomes 2000. *Cytogenetics and Cell Genetics* **90**: 169–218.

Schnabel, R. D., Kim, J.-J., Ashwell, M. S., Sonstegard, T. S., Van Tassell, C. P., Connor, E. E. and Taylor, J. F. (2005). Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 6896–6901.

Schwaiger, F. W., Gostomski, D., Stear, M. J., Duncan, Q. A., J. L.and McKeller, Epplen, J. T. and Buitkamp, J. (1995). An ovine *Major histocompatibility complex DRB1* allele is associated with low fecal egg counts following natural, predominantly *Osteragia circumcincta* infection. *International Journal for Parasitology* **25**: 815–822.

Schwartz, M. K., Tallmon, D. A. and Luikart, G. (1998). Review of DNA-based census and effective population size estimators. *Animal Conservation* **1**: 293–299.

Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M. and Visscher, P. M. (2002). QTL Express: Mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* **18**: 339–340.

Segers, K., Vaiman, D., Berghmans, S., Shay, T., Meyers, S., Beever, J., Cockett, N., Georges, M. and Charlier, C. (2000). Construction and characterization of an ovine BAC contig spanning the callipyge locus. *Animal Genetics* **31**: 352–359.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**: 605–610.

Simm, G. (1998). *Genetic Improvement of Cattle and Sheep.* Farming Press, Ipswich.

Sinsheimer, J. S., Blangero, J. and Lange, K. (2000). Gamete-competition models. *American Journal of Human Genetics* **66**: 1168–1172.

Slate, J., Coltman, D. W., Goodman, S. J., MacLean, I., Pemberton, J. M. and Williams, J. L. (1998). Bovine microsatellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*) and Soay sheep (*Ovis aries*). *Animal Genetics* **29**: 307–315.

Slate, J., Pemberton, J. M. and Visscher, P. M. (1999). Power to detect QTL in a free-living polygynous population. *Heredity* **83**: 327–336.

Slate, J., Van Stijn, T. C., Anderson, R. M., McEwan, K. M., Maqbool, N. J., Mathias, H. C., Bixley, M. J., Stevens, D. R., Molenaar, A. J., Beever, J. E., Galloway, S. M. and Tate, M. L. (2002a). A deer (subfamily Cervinae) genetic linkage map and the evolution of ruminant genomes. *Genetics* **160**: 1587–1597.

Slate, J., Visscher, P. M., MacGregor, S., Stevens, D., Tate, M. L. and Pemberton, J. M. (2002b). A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). *Genetics* **162**: 1863–1873.

Smit, M., Segers, K., Carrascosa, L. G., Shay, T., Baraldi, F., Gyapay, G., Snowder, G., Georges, M., Cockett, N. and Charlier, C. (2003). Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* **163**: 453–456.

Soule, M. E., ed. (1986). *Conservation Biology: The Science of Scarcity and Diversity.* Sinauer and Associates, Sunderland, MA.

Souza, C. J. H., MacDougall, C., Campbell, B. K., McNeilly, A. S. and Baird, D. T. (2001). The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *Journal of Endocrinology* **169**: R1–R6.

Sponenberg, D. (1997). Genetics of colour and hair texture. In Piper, L. and Ruvinsky, A., eds., *The Genetics of Sheep*, page 5186, CAB International, Cambridge.

Stevenson, I. R., Marrow, P., Preston, B. T., Pemberrton, J. M. and Wilson, K. (2004). Adaptive reproductive strategies. In Clutton-Brock, T. and Pemberton, J., eds., *Soay Sheep: Dynamics and Selection in an Island Population*, pages 243–275, Cambridge University Press, Cambridge.

Stone, R. T., Keele, J. W., Shackelford, S. D., Kappes, S. M. and Koohmaraie, M. (1999). A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *Journal of Animal Science* **77**: 1379–1384.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**: 1171–1177.

Taillon-Miller, P., Bauer-Sardiña, I., Saccone, N. L., Putzel, J., Laitinen, T., Cao, A., Kere, J., Pilia, G., Rice, J. P. and Kwok, P.-Y. (2000). Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nature Genetics* **25**: 324–328.

Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K., Furugaki, K., Matsumoto, T., Mano, S., Ando, S., Nozaki, Y., Yukawa, W., Nakashige, R., Yamaguchi, D., Ishibashi, H., Yonekura, M., Nakami, Y., Takayama, S., Endo, T., Saruwatari, T., Yagura, M., Yoshikawa, Y., Fujimoto, K., Oka, A., Chiku, S., Linsen, S. E. V., Giphart, M. J., Kulski, J. K., Fukazawa, T., Hashimoto, H., Kimura, M., Hoshina, Y., Suzuki, Y., Hotta, T., Mochida, J., Minezaki, T., Komai, K., Shiozawa, S., Taniguchi, A., Yamanaka, H., Kamatani, N., Gojobori, T., Bahram, S. and Inoko, H. (2005). Whole genome association study of rheumatiod arthritis using 27039 microsatellites. *Human Molecular Genetics* **14**: 2305–2321.

Tenesa, A., Knott, S. A., Ward, D., Smith, D., Williams, J. L. and Visscher, P. M. (2003). Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science* **81**: 617–623.

Terwilliger, J. D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *American Journal of Human Genetics* **56**: 777–787.

Terwilliger, J. D. and Weiss, K. M. (1998). Linkage disequilibrium mapping of complex disease: Fantasy or reality? *Current Opinion in Biotechnology* **9**: 578–594.

Thomas, P. D. and Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidece for differences in molecular effects. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 15398–15403.

Thompson, J. M. and Ball, A. J. (1997). Genetics of meat quality. In Piper, L. and Ruvinsky, A., eds., *The Genetics of Sheep*, pages 523–538, CAB International, Wallingford.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach.* Clarendon Press, Oxford.

Tranulis, M. A. (2002). Influence of the prion protein genem *Prnp*, on scrapie susceptibility in sheep. *APMIS* **110**: 33–43.

Vaiman, D., Schibler, L., Bourgeois, F., Oustry, A., Amigues, Y. and Cribiu, E. P. (1996). A genetic linkage map of the male goat genome. *Genetics* **144**: 279–305.

Van Laere, A.-S., Nguen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., Andersson, G., Georges, M. and Andersson, L. (2003). A regulatory mutation in *IGF2* causes a major QTL growth effect on muscle growth in the pig. *Nature* **425**: 832–836.

Van Ooijen, J. W. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics* **84**: 803–811.

Visscher, P. M., Haley, C. S., Heath, S. C., Muir, W. J. and Blackwood, D. H. R. (1999). Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatric Genetics* **9**: 75–84.

Visscher, P. M., Thompson, R. and Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**: 1013–1020.

Wall, J. D. and Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* **4**: 587–597.

Walling, G. A., Visscher, P. M., Simm, G. and Bishop, S. C. (2001). Confirmed linkage for QTLs affecting muscling in Texel sheep on Chromosome 2 and 18. In *52nd Annual Meeting of the European Association for Animal Production*, Budapest.

Walling, G. A., Visscher, P. M., Wilson, A. D., McTeir, B. L., Simm, G. and Bishop, S. C. (2004). Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. *Journal of Animal Science* **82**: 2234–2245.

Wan, Y., Cohen, J. and Guerra, R. (1997). A permutation test for the robust sib-pair linkage method. *Annals of Human Genetics* **61**: 79–87.

Wang, J. and Caballero, A. (1999). Developments in predicting the effective population size of subdivided populations. *Heredity* **82**: 212–226.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.

Weller, J. I., Kashi, Y. and Soller, M. (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**: 2525–2537.

Williams, J. L. (2005). The use of marker-assisted selection in animal breeding and boitechnology. *Revue Scientifique et Technique de l Office International des Epizooties* **24**: 379–391.

Williams, J. L., Eggen, A., Ferretti, L., Farr, C. J., Gautier, M., Amati, G., Ball, G., Caramorr, T., Critcher, R., Costa, S., Hextall, P., Hills, D., Jeulin, A., Kiguwa, S. L., Ross, O., Smith, A. L., Saunier, K., Urquhart, B. and Waddington, D. (2002). A bovine whole-genome radiation hybrid panel and outline map. *Mammalian Genome* **13**: 469–474.

Williams, J. T. and Blangero, J. (1999). Power of variance component linkage analysis to detect quantitative trait loci. *Annals of Human Genetics* **63**: 545–563.

Williams, J. T., Duggirala, R. and Blangero, J. (1997). Statistical properties of a variance components method for quantitative trait linkage analysis in nuclear families and extended pedigrees. *Genetic Epidemiology* **14**: 1065–1070.

Wilson, J. F. and Goldstein, D. B. (2000). Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *American Journal of Human Genetics* **67**: 926–935.

Wilson, T., Wu, X. Y., Juengel, J. L., Ross, I. K., Lumsden, J. M., Lord, E. A., Dodds, K. G., Walling, G. A., McEwan, J. C., O'Connell, A. R., McNatty, K. P. and Montgomery, G. W. (2001). Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction* **64**: 1225–1235.

Woolaston, R. R. and Piper, L. R. (1996). Selection of Merino sheep for resistance to *Haemonchus contortus*: Genetic variation. *Animal Science* **62**: 451–460.

Wray, N. R., Woolliams, J. A. and Thompson, R. (1990). Methods for predicting rates of inbreeding in selected populations. *Theoretical and Applied Genetics* **80**: 503–512.

Wright, S. (1922). Coefficient of inbreeding and relationship. *American Naturalist* **56**: 330–338.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Xu, S. and Atchley, W. R. (1995). A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**: 1189–1197.

Zeng, Z.-B., Kao, C.-H. and Basten, C. J. (1999). Estimating the genetic architecture of quantitative traits. *Genetical Research* **74**: 279–289.

Zou, F., Yandell, B. S. and Fine, J. B. (2001). Statistical issues in the analysis of quantitative traits in combined crosses. *Genetics* **158**: 1339–1346.