Genetic variation in natural populations: linkage mapping of fitness related traits in Soay sheep

Dario Beraldi

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2006

Don Quixote's disgrace is not his fantasy, but Sancho Panza.

Franz Kafka

Do not despair: one of the thieves was saved. Do not presume: one of the thieves was damned.

Saint Augustine

"One of the thieves was saved. *(Pause.)* It's a reasonable percentage".

Samuel Beckett, Waiting for Godot

Table of contents

Acknowledgments

I thank J.G. Pilkington, B.H. Craig and many previous volunteers and project members for collecting field data and genetic samples, and A.J. Wilson for statistical advice. I thank the National Trust for Scotland for granting permission to work on St. Kilda, and QinetiQ for logistical support. I am also grateful to the Queensland Institute of Medical Research for providing the facilities to perform part of the statistical analysis. The long term data collection on St. Kilda has been supported by Natural Environment Research Council (NERC) and Wellcome Trust grants to T.H. Clutton-Brock, B.T. Grenfell, L.E.B Kruuk, M.J. Crawley and J.M. Pemberton. This study was funded by NERC through its Environmental Genomics thematic programme (grant number NER/T/S/2002/00189).

There are a number of friends whose moral support during my PhD project I need to acknowledge. In a very incomplete list, these are my laboratory and office mates Silvia Perez Espona in particular, my pal Ronnie Mooney, Angeles de Cara, and my flatmates Juan Carlos Ruiz Guajardo and Sonja Preuss. Without them my work would have been much less exciting and, perhaps, not even possible. My supervisors (Josephine Pemberton and Peter Visscher) and colleagues (Jon Slate, Jake Gratten and Allan McRae) have been great not only for their scientific advice but also for their encouragement. Finally, I thank Andrea Mazzucato for the enthusiasm in scientific research he gave me.

List of publications

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M., Pemberton, J.M., 2006. Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (*Ovis aries*). Genetics 173, 1521-1537.

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M., Pemberton, J.M. Quantitative trait loci (QTL) mapping of resistance to strongyle and coccidia in the free-living Soay sheep (*Ovis aries*). Int J. Parasitol. In press.

McRae, A.F., **Beraldi, D.,** 2006. Examination of a region showing linkage map discrepancies across sheep breeds. Mamm. Genome 17, 346-353.

Csilléry, K., Johnson, T., **Beraldi, D.,** Clutton-Brock, T.H., Coltman, D., Hansson, B., Spong, G., Pemberton, J.M., 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics 173, 2091-2101.

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M., Pemberton, J.M. Mapping QTL underlying fitness-related traits in a free-living sheep population. Submitted to Evolution.

Gratten J., **Beraldi D.,** Lowder B.V., McRae A.F., Visscher P.M., Pemberton P.M., Slate J. A single nucleotide substitution in TYRP1 is responsible for dark-light coat colour variation in a free-living population of Soay sheep. Proc. R. Soc. Lond. B Biol. Sci. In press.

Abstract

Traits under the pressure of natural selection are expected to lose their genetic variability since the negative genetic variants are counter-selected and replaced by the positive alleles. However, direct, empirical observations show that fitness related traits maintain, or even enhance, their genetic variability in the face of selection. Explaining the mechanisms of the maintenance of genetic variation is of fundamental importance in understanding the process of evolution in nature. One strategy by which the complexity of continuous characters can be dissected and the loci that contribute to trait variation identified is linkage mapping. Applied within the natural environment, linkage mapping can potentially shed light on the evolutionary mechanisms behind the selection and maintenance of genetic diversity in wild populations. Thanks to previous studies, the Soay sheep on St. Kilda, the free-living population which is the subject of this project, offer an unprecedented opportunity to dissect the architecture of fitness related traits. In this project, 588 animals belonging to a larger pedigree (882 animals) were genotyped to build a complete genetic map comprising 251 markers. Subsequently, a whole genome scan was carried out to detect qualitative and quantitative trait loci. Different morphological and physiological traits were chosen on the basis of their relationship to fitness. The loci affecting the Mendelian polymorphisms of horn type and two aspects of coat colour were mapped with good reliability (*Horn type* mapped to chromosome 10, *Coat colour* to chromosome 2, *Coat pattern* to chromosome 13; LOD scores >3). Several quantitative traits were analyzed and one genomic region showed significant linkage with the variation in jaw length (chromosome 11, LOD= 3.59). In addition, suggestive linkages were detected for hind leg length (chromosome 15, LOD= 2.89), birth weight (chromosome 8, LOD= 2.54), and timing of birth (chromosome 2, LOD= 2.70). Finally, two suggestive linkages were scored for resistance to the gastrointestinal parasite coccidia

(chromosome 3 and X, LOD= 2.68 and 2.21, respectively). As the quantitative trait loci mapping was performed by variance component analysis, the parameters describing variation in the study traits were also estimated. This project is one of the first attempts to dissect the genetic architecture of complex traits in the wild through linkage mapping. The results reported here improve our understanding of the evolution and genetics of natural populations by allowing more realistic modelling of the study traits and by opening the way towards the molecular characterization of the genomic regions of interest.

Chapter 1.

Introduction to QTL mapping in natural populations

1.1. Background

The paradox of genetic variation – The maintenance of genetic variation in traits under natural selection is one of the most important and challenging problems in evolutionary biology. Fisher's Fundamental Theorem of Natural Selection (Fisher, 1958) states that "the rate of increase in fitness of an organism at any time is equal to its genetic variance in fitness at that time". A predicted consequence is that additive genetic variation is depleted by selection, so that traits related to fitness tend to have low or null values of heritability since the alleles conferring higher fitness should reach fixation (Falconer, 1989). To test this principle, Mousseau and Roff (1987) reviewed a large dataset of estimates of heritability for traits under selection collected from wild populations. Their analysis showed that, on the one hand, life history traits under selective pressure had lower heritability than morphological traits, thus supporting Fisher's theorem, but, on the other hand, even genes for traits under strong selection were not fixed as was expected. Furthermore, when the additive genetic variance was standardized by the trait mean (giving the coefficient of variation) instead of by the total phenotypic variance (giving the heritability), traits affecting fitness had *more* additive genetic variation than the morphological ones (Houle, 1992).

A number of different, although not mutually exclusive theories have been suggested to explain the maintenance of genetic variation under selection (Barton and Keightley,

2002). For example, the phenomenon of negative pleiotropy may prevent fixation as alleles selected for in one trait are also selected against in another trait (Zhang and Hill, 2002). Epistasis, gene by environment interactions, and balancing selection may also counteract fixation (Hill, 1982; Kondrashov and Turelli, 1992; Carlborg and Haley, 2004; Carlborg et al., 2006). Furthermore, natural populations can inhabit a spatially heterogeneous and changing environment which means that the fittest genetic combination is elusive (Olendorf et al., 2006). The relative prevalence and weight of these phenomena in explaining natural variation is still a matter of debate and few studies have been conducted to test the different hypotheses (notable exceptions are Kroymann and Mitchell-Olds, 2005; Olendorf et al., 2006). Ideally, a comprehensive quantitative genetics theory of evolution would allow the characterization of the forces maintaining and depleting variation. This issue is relevant not only from a purely evolutionary research point of view, but also for practical progress in medicine and agriculture (Risch, 2000; Andersson, 2001) as many diseases and economic traits have a quantitative basis shaped by the laws of natural selection.

Linkage mapping in experimental and natural populations – Linkage analysis investigates the correlation between phenotypic values and one or more genetic markers to find genomic regions that influence the phenotype. A genetic marker may be thought of as an evolutionary neutral DNA sequence which is variable across individuals in a population and whose position in the genome is known. A large correlation between phenotypic value and marker genotype suggests that the tested marker is close to a region affecting the trait. By genotyping each member of a family or pedigree, one can trace the inheritance pattern at each marker locus. Mapping quantitative trait loci (QTL) is a strategy to dissect the complexity of continuous traits, such as size, fitness and disease susceptibility (Lynch and Walsh, 1998; Glazier et al., 2002), and so contributes to the

formulation of better models to describe the nature of quantitative variation (Erickson et al., 2004; Erickson, 2005; Slate, 2005). Moreover, coupled with linkage disequilibrium mapping and comparative analysis, linkage mapping may also open the way to isolating and characterizing the genes underlying variation at the molecular level (Frary et al., 2000; Van Laere et al., 2003).

QTL mapping projects are now routinely carried out in model organisms (Mackay, 2001) or in plant and animal species of agricultural relevance (Evans et al., 2003; Khatkar et al., 2004; Mackay, 2004). These studies can rely on experimental crossings and environments to maximize the power of statistical analysis by reducing the sources of variation other than the genetic one. The results obtained therefore have good confidence and accuracy but they cannot be readily transferred to natural populations since differences between natural and artificial frameworks exist both in the environment and in the population itself. Experimental populations are often highly or completely inbred so that the offspring derived from crossing divergent lines have complete linkage disequilibrium in the first generation (F_1) and consequently the following F_2 or BC_1 progenies become informative (i.e. segregating) at all markers and genes differing between parental lines. Further, to increase the number of polymorphic loci and the contrast in trait means, the study populations are sometimes derived from individuals belonging to different species. QTLs detected in this way represent differences between parental lines and therefore the findings obtained do not contribute directly to the understanding of the variability within populations. QTL mapping in outbred livestock investigates variation within populations and are therefore closer to natural populations.

Although the findings discovered with outbred crosses are expected to have more general validity, outbred populations are inherently less powerful than inbred lines to map QTL. Compared to inbred lines, crosses between outbred individuals have fewer

informative meioses since not all the families segregate at the marker loci and QTLs. This is due to the fact that the parents employed in outbred crosses are genetically variable, whereas in inbred line crosses parents are genetically uniform. Also, the marker and QTL linkage phase may not be conserved across families thus requiring appropriate statistical solutions (Lynch and Walsh 1998). QTL effects are expressed as means with inbred lines but as variances with outbred lines, thus resulting in a less accurate QTL effect estimate from outbred populations (Lynch and Walsh 1998). Despite the fact that outbred crosses are closer to natural populations than inbred ones, they often undergo genetic drift and selection within the laboratory or farm, which simplifies and reduces the population structure and variability. As a result, outbred crosses may lose pleiotropic and epistatic effects that can play an important role in natural environments (Carlborg and Haley, 2004; Kroymann and Mitchell-Olds, 2005). For these reasons the genetics of outbred populations may not be representative of unmanaged pedigrees or families collected directly from the wild to study natural variation.

The environment plays a fundamental role in determining the phenotype as the genetic make-up of an organism represents *only* the potential for a trait to develop in response to a given environment. In terms of environment, the natural environment is not constant and predictable across time and space in contrast to the experimental setting. This heterogeneity can explain, at least in part, the maintenance of genetic variation. In some cases, in fact, different genotypes respond in nonparallel ways to different environments (genotype by environment interaction). The genotype by environment interaction leads to the development of locally adapted strains and therefore it can be regarded as a key component of speciation and evolution, apart from also having important consequences in breeding and genetic improvement. Laboratory or agricultural experiments are often designed to enhance the properties of the trait under

study by minimizing the sources of environmental variance so that the genes segregating in the wild may be not the same of those identified in captivity and the estimates of heritability may become inflated (Houle, 1992). Despite the interest in closing the gap between natural and experimental studies, few projects have been undertaken so far to map QTL in the wild (Slate et al., 2002; Erickson et al., 2004; Erickson, 2005; Slate, 2005). An exception could be in humans where, however, cultural and ethic factors make extension of findings to animal populations inappropriate, and it could be questionable whether selection in the human population is natural and applicable to animal populations. The paucity of studies of the genetic architecture of complex traits in the wild is due to either the lack of suitable populations, especially in terms of comprehensive phenotypic datasets and pedigrees, or to the difficulties in statistical analysis behind such studies (Hoeschele et al., 1997; Flint and Mott, 2001).

A recently developed statistical tool, namely variance component analysis, is particularly suitable for the analysis of wild populations. Variance component estimation by restricted maximum likelihood (REML) exploits the information derived from any pair of individuals in the data-file thus making optimal use of the information contained in large, extended pedigrees (Hoeschele et al., 1997; Visscher et al., 1999). Coupled with Markov chain Monte Carlo algorithms, variance component QTL analysis has the power to accommodate pedigrees of any size and complexity. In addition, deviations from the assumption of normality of the distribution of data and unbalanced or missing data generally do not dramatically bias the parameter estimation (Allison et al., 1999; Kruuk, 2004). Finally, provided that the dataset is large enough, gene by gene (epistatic) or genotype by environment interactions can be modelled and estimated. This method has been largely adopted and developed by human geneticists (Almasy and Blangero 1998) and recently applied to animal breeding and wild populations as an extension of the

animal model (George et al., 2000; Slate et al., 2002). The principle (Lynch and Walsh, 1998) is to use marker information to infer, for any genomic region of interest, the genetic fraction that is identical by descent between two individuals. One reason for considering the number of alleles identical by descent instead of by state is that the distribution of the former only depends on the relative type, whereas the latter also depends on the allele frequencies. A maximum likelihood function associated with a putative QTL is then modelled only on a target region without the need for modelling all the underlying genetic details. In general, the phenotypic value, assumed to follow a multivariate normal distribution, is partitioned into a random effect due to a major gene (QTL), a polygenic effect, and an error. Additional random effects such as maternal and permanent environmental effect can be accounted for. The covariance between the trait values of two individuals is then a function of the chromosomal region identical by descent, which accounts for the additive genetic variance explained by a putative major gene, and the coefficient of coancestry, which accounts for genetic variance not explained by the QTL. The presence of a QTL in a given region is tested by a likelihood-ratio test comparing a model where the additive genetic variance of the QTL is set to zero against a model where this variance is greater than zero.

1.2. The St. Kilda islands and the Soay sheep

The most important reason why few QTL mapping projects have been carried out in the wild so far is, probably, the lack of free-living populations where phenotypic and genetic data have been collected. In this respect, the free-living Soay sheep population on St. Kilda stands out as an exception since they are unmanaged but well studied animals. A comprehensive description of the Soay sheep as a study population for genetics and evolution can be found in Clutton-Brock and Pemberton (2004) and references therein.

The St. Kilda archipelago – Lying about 160 km west of mainland Scotland, St. Kilda is one of the most remote British islands (Figure 1.1). The erosion of an ancient volcano gave rise to St. Kilda, an archipelago of four main islands and several stacks. The largest and also most accessible island is Hirta, 637 ha, while Soay to the north-west and Dun to the south comprise 99 and 32 ha respectively. Seven km to the north west of Hirta is Boreray (77 ha) together with two large stacks.

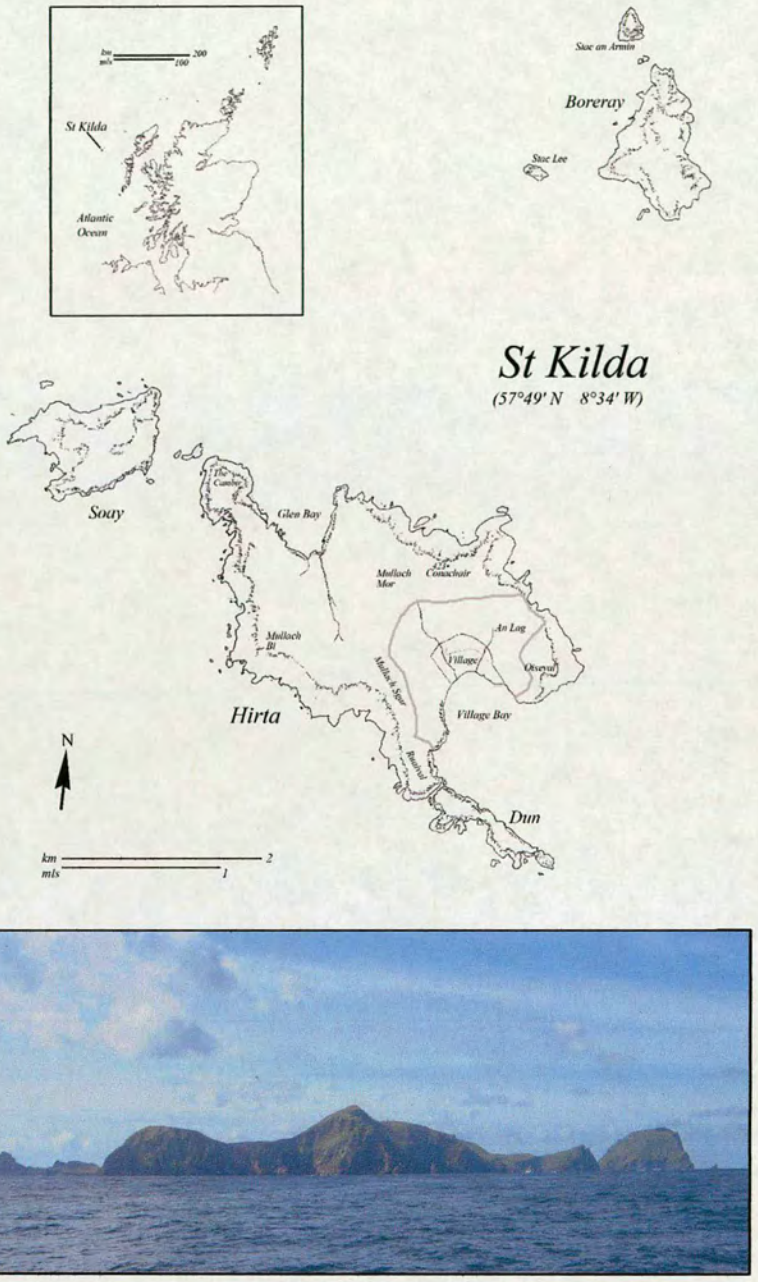

Figure 1.1. – The St. Kilda archipelago: geographical position (image from Ian Stevenson's PhD thesis, 1994) and panoramic view.

St. Kilda is an excellent site for wildlife. Among other species it is home to a remarkable colony of sea birds and an endemic sub-species of field mouse (*Apodemus sylvaticus*), this latter was probably introduced with human occupation and evolved into a form of unique natural interest. The naturalistic importance of St. Kilda was

acknowledged in 1957 when it became a National Nature Reserve property of the National Trust for Scotland (Boyd et al., 1964) and in 1986 when UNESCO designated St. Kilda World Heritage Site.

Bronze Age remains and other archaeological evidence show that the islands of St. Kilda have been inhabited more or less continuously for 2000 years. However, in 1930, after a period of population decline begun in the 18th century, the 36 remaining St. Kildans were evacuated due to their inability to collect enough food (Steel, 1988). The human impact was largely limited to the construction of cleits (dry stone food storage buildings) and other structures and the cutting of turf for fuel (long since ceased) together with the exploitation of the seabirds for food. Notably, despite such a long history of human occupation, there have been few alien species introductions, mostly a few plants within a confined location.

The Soay sheep population – The Soay sheep (Figure 1.2) owes its name to the island of Soay where it is thought that the Vikings introduced these animals during the 9th or 10th century. The Soay sheep can be regarded as the most primitive breed in Europe, resembling in different features their wild ancestors, the mouflon, and the archaeological remains of the early Neolithic age (Campbell, 1974; Doney et al., 1974). During their isolation, Soay sheep remained mostly unchanged although the possibility cannot be excluded that the inhabitants of Hirta attempted to improve the Soay flock by means of mating with domestic sheep (Boyd et al., 1964). After the evacuation of the St. Kildans from Hirta, the domestic animals that were not caught were shot and in order to manage the vegetation, 107 sheep from Soay were transferred to Hirta. Nowadays, the sheep population on Hirta numbers between 600 and 2000 individuals. This figure is highly variable since the Soay sheep experience population fluctuations consisting of winter crashes when up to 70% of the animals die (Coulson et al., 2001; Clutton-Brock et al.,

2004a). Summers are characterized by high levels of food and fecundity which after 2-3 years bring the population to a level higher than the island can support over winter so that a following rainy and windy winter causes the population to crash.

Despite the small size of the island and the absence of restrictions to the movement of the sheep, the animals living in the Village Bay area do not constitute a single randomly mated population. Demographic studies revealed three sub-populations that are characterized by different survival, recruitment and dispersal rates (Coulson et al., 1999). These three groups are also genetically distinct as shown by the analysis of the F_{ST} and F_{IS} through microsatellite polymorphism (Coltman et al., 2003). This population structure is driven by the incomplete postnatal dispersal of females which, in contrast to males, tend to share the same location as their mothers.

Figure 1.2 – Party of three Soay rams (photograph by Andrew MacColl)

Research expeditions – Thanks to the isolation on St. Kilda, absence of natural predators, and characteristic population dynamics, the Soay sheep lend themselves to evolutionary and ecological studies in a wild but relatively simple framework. For this reason extensive efforts have been made to collect data about them. Since 1985, regular expeditions have been sent to St. Kilda to record and monitor the population dynamics.

All sheep living in Village Bay area (175 ha), where about one third of the Hirta population lives, are tagged as far as possible. Regular censuses are performed throughout the Village Bay study area in order to identify each sheep and to record where it is and which other sheep it is with. The first expedition of the year occurs in April and early May when new born lambs are ear tagged and weighed, and a tissue sample is taken for later genetic analysis. At this stage it is also possible to assign the maternity and date of birth of each individual. The summer expedition is aimed at catching as many individuals as possible; usually, 65% of the animals within Village Bay are caught and information of interest collected. The traits that are recorded for each animal are: body weight to the nearest 0.1 kg, hind length leg to the nearest millimetre, horn type and length, testicular circumference. In addition, a faecal sample is taken to estimate the degree of parasite infection, and a blood sample is collected for subsequent genetic analyses. Some individuals (56% of females and 25% of males) have been caught more than once thus providing repeated measures for the same animal. Finally, a third expedition from mid-October to the end of November is carried out to study the rut, and to capture males that immigrate to the study area to mate. The main repository for the information so gathered is the *Soay Sheep Database,* a Microsoft Access file implemented by Sunadal Data Solutions (Edinburgh) and maintained by the Soay Sheep Project members. Currently, phenotypic data for more 6000 sheep is stored in the database.

Parentage assignment – Maternal relationships are assigned by field observation and confirmed by molecular analysis so that the degree of accuracy is high. Paternal inference instead relies mainly on molecular analysis since during the rutting season dams may mate with multiple candidate sires and therefore field observations of mating are not reliable (Coltman, 2005). All sampled individuals are genotyped at up to twenty unlinked

microsatellite or allozyme loci (Overall et al., 2005). The genotype so derived are analyzed by CERVUS (Marshall et al., 1998), a likelihood-based program specifically designed for parentage assignment in natural populations. CERVUS employs a likelihood procedure to identify the most likely fathers for a given offspring. The final assignment is then resolved by a simulation algorithm to discriminate between the most likely and the next most-likely father. CERVUS takes into account the possibility that genotyping error may occur with a certain frequency, that only a proportion of the candidate fathers is sampled, and that not all the individuals are typed at all the loci. These shortcomings are relatively common in large scale studies of wild populations. At the moment, the Soay sheep pedigree numbers more 3000 sheep.

1.3. Aim and overview of the project

As an attempt to gain insights into the genetic architecture of complex traits in the wild, this project aims to map qualitative and quantitative trait loci affecting morphological and parasite resistance traits related to fitness in Soay sheep by taking advantage of the wealth of information gathered so far on this unmanaged population. Provided that the target traits are heritable and major genes exist, the Soay sheep have the pre-requisites to allow linkage mapping: as described above, pedigree and phenotypic data are available and many molecular markers (microsatellites) have been identified in domestic sheep (Maddox et al., 2001).

In order to carry out a genome scan it is necessary to make use of a number of related individuals which have the relevant phenotypic data known. In addition, the selected individuals need to be genotyped at several marker loci in order to build a linkage map covering, ideally, the whole genome. This material and information will then be combined in a statistical framework to find an association between a trait value and

specific genetic region. To this end, chapter 2 describes the selection of a mapping panel (*i.e.* the related individuals to be genotyped) and the construction of the Soay sheep linkage map. The map comprised 247 microsatellite and four allozyme markers and covers approximately 3350 cM (~90% of the sheep genome) with 15 cM inter-marker interval. The linkage mapping of three discrete traits of evolutionary interest (coat colour, coat pattern, and horn type) is also described. The *Coat colour* locus maps to chromosome 2 where a strong candidate gene, tyrosinase-related protein 1 (TYRP1), has also been mapped. *Coat pattern* maps to chromosome 13, close to the candidate locus *Agouti*. *Horn type* maps to chromosome 10, a location similar to that previously identified in domestic sheep (Montgomery et al., 1996). These findings represent an advance in the dissection of genetic diversity in the wild and provide the foundation for the subsequent QTL analyses.

In chapter 3, the linkage map and mapping panel are used to detect QTL responsible of a variety of morphometric and early developmental traits related to the individual fitness. In this chapter, the use of variance component analysis for QTL mapping in a free-living population is also described. The study traits include birth date and weight, considered both as maternal and offspring traits, fore and hind leg length, body weight, jaw length, and metacarpal length. Genetic and environmental components of phenotypic variance were estimated for each trait and, for those traits showing non-zero heritability, a QTL search was conducted by variance component analysis. Support for a QTL at genome-wide significance was found on chromosome 11 for jaw length; suggestive QTL were found on chromosomes 2 (for birth date as a trait of the lamb), 8 (birth weight as a trait of the lamb) and 15 (adult hindleg length). Here are also discussed the prospects for refining estimates of QTL position and effect size in the study population, and for QTL searches in free-living pedigrees in general.

In a fashion similar to chapter 3, chapter 4 reports the use of variance component analysis to detect QTL affecting resistance to gastrointestinal parasites and ectoparasitic keds (*Melophagus ovinus*). The traits here investigated were the strongyle faecal egg count, the coccidia oocyst count, and the count of keds. Two genomic regions reached the level of suggestive linkage for coccidia oocyst count in lambs (LOD= 2.68 and 2.21 on chromosome 3 and X, respectively). This is the first study to report a QTL search for parasite resistance in a free-living animal population and therefore may represent a useful reference for similar studies aimed at understanding the genetics of host-parasite co-evolution in the wild.

Finally, the potential future directions of the present work are discussed with respect to the improvement of the current results in terms of accuracy and reliability and with respect to the possible use of the results for follow-up studies. To this end, the map developed here has facilitated the fine mapping of the *Coat colour* locus (Gratten et al. In press), the comparison with other sheep maps (McRae and Beraldi, 2006), and the evaluation of relatedness estimator precision in natural populations (Csilléry et al., 2006).

***

Chapters 2, 3 and 4 are presented in form of stand-alone manuscripts as they have been, or they are about to be, submitted as research articles to appropriate scientific journals (see List of publications on page 6).

Chapter 2.

Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (*Ovis aries*).

Dario Beraldi[1,4], Allan F. McRae[1,2,5], Jacob Gratten[3,6], Jon Slate[3,7], Peter M. Visscher[1,2,8] and Josephine M. Pemberton[1,9]

[1] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, UK.*

[2] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, 4029, Australia.*

[3] *Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK.*

[4] dario.beraldi@ed.ac.uk; [5] allan.mcrae@qimr.edu.au; [6] j.gratten@sheffield.ac.uk; [7] j.slate@sheffield.ac.uk; [8] peter.visscher@qimr.edu.au; [9] j.pemberton@ed.ac.uk

Author contribution: DB carried out the laboratory work, performed the statistical analysis, and wrote the manuscript. AFM and PMV helped with the statistical analysis. All the contributing authors critically read and improved drafts of this manuscript. The whole project was planned and supervised by JMP, JS and PMV.

2.1. Abstract

An understanding of the determinants of trait variation and the selective forces acting on it in natural populations would give insights to the process of evolution. The combination of long term studies of individuals living in the wild and better genomic resources for non-model organisms makes achieving this goal feasible. This manuscript reports the development of a complete linkage map in a pedigree of free-living Soay sheep on St. Kilda, and its application to map the loci responsible for three morphological polymorphisms for which the maintenance of variation demands explanation. The map was derived from 247 microsatellite and four allozyme markers and covers 3350 cM (approximately 90% of the sheep genome) at approximately 15 cM intervals. Marker order was consistent with the published sheep map with the exception of one region on chromosome 1 and one on chromosome 12. *Coat colour* maps to chromosome 2 where a strong candidate gene, Tyrosinase Related Protein 1 (TYRP1) has also been mapped. *Coat pattern* maps to chromosome 13, close to the candidate locus *Agouti*. *Horn type* maps to chromosome 10, a similar location to that previously identified in domestic sheep. These findings represent an advance in the dissection of the genetic diversity in the wild, and provide the foundation for QTL analyses in the study population.

Keywords: linkage mapping, wild population, coat colour, horn type, genetic map, Soay sheep

2.2. Introduction

An area of fundamental research in evolutionary genetics concerns the closely related issues of understanding the determinants of trait variation in natural populations and understanding how genetic variation for traits is maintained in the face of natural selection. The first of these problems is often summarised as the 'genetic architecture' question: in general we would like to know whether genes of large effect commonly segregate in natural populations, or whether the infinitesimal model, i.e. that most traits are controlled by many genes of small effect, is appropriate – or perhaps more likely, some distribution in between (Barton and Keightley, 2002; Brem and Kruglyak, 2005). Similarly, we would like to know to what extent genetic interactions such as dominance, pleiotropy and epistasis contribute to the evolutionary dynamics of a population. The second problem was long ago identified by Fisher (Fisher, 1958): how is it that genetic variation for traits persists when selection is so often directional? The answer to this question must lie not only in the genetic architecture question, but also in the modes of selection and their temporal and spatial stability.

In principle, the arrival of abundant molecular markers, genetic maps and whole genome sequences allows us to address both genetic architecture and selection in much greater depth than ever before, since the role of variation at individual loci can be assessed. Mapping trait loci is a starting point to providing information on the genetic architecture of a trait in terms of the number of genes involved, relative effect, and mode of expression (Erickson et al., 2004; Slate, 2005). In turn, this allows study of the relationship between phenotype and genotype, and inference of the selective forces acting on the critical locus. Furthermore, by mapping genes, it is possible to test for the presence of gene-by-gene (epistatic) and gene-by-environment interactions, which are thought to contribute to phenotypic variation in natural and controlled settings (Carlborg

and Haley, 2004; Erickson, 2005). In addition, the discovery of the map location of genes that influence phenotypic variation means that patterns of linkage disequilibrium (LD) and haplotype structure can be examined, which may provide insights about population history and selection Unfortunately, some of the characteristics that make experimental populations so practical for linkage mapping also restrict the degree to which findings can be extrapolated to natural populations. Usually, geneticists generate segregating populations derived from one or a few pair of parents which are often inbred and selected for the extreme phenotypes. In addition, the population is raised in a uniform and controlled environment (e.g. a greenhouse) where non-genetic sources of phenotypic variation are minimised. On the one hand, this strategy maximizes the power of analysis, i.e. it increases the probability of finding a statistical association between marker genotype and phenotypic trait, but on the other hand, as the aim of genetic research becomes the understanding of how selection shapes genomes, the findings in experimental crosses are of limited applicability (Roff and Simons, 1997; Conner, 2002). In the wild, individuals are exposed to environmental and genetic forces (e.g. genotype-by-environment interaction, pleiotropy, epistasis, maternal effects) some of which are unwittingly or deliberately diminished in experimental settings, and may conceal important effects in the wild (Kroymann and Mitchell-Olds, 2005; Wilson et al., 2005a; Wilson et al., 2005b). Although these forces are particularly difficult to detect in the wild, their possible absence from an experimental design may lead to biased conclusions.

A refined understanding of the process of evolution can be expected if the precise loci underlying trait variation can be identified and their behaviour studied in free-living populations. Hence, a recent development is the application of genomic analyses to studies of free-living populations. Techniques for generating large numbers of genetic markers (e.g. AFLPs and microsatellites), and the availability of markers from related

model species means that genetic maps and Quantitative Trait Locus (QTL) searches in organisms originally studied in the wild are becoming more common (Erickson et al., 2004; Slate, 2005). To date, most of these studies have involved wild plants or animals brought into and bred in the laboratory. Although in some cases the experimental design makes use of pedigrees generated from several lines (Zhang et al., 2005) and investigate fitness related traits (Laurie et al., 2004), such studies do not directly address the action of natural selection as the study organisms are the product of breeding programs. Other projects have been designed to answer specific ecological or evolutionary questions, and to this end have employed individuals drawn from the wild and crossed under controlled conditions (Hawthorne and Via, 2001; Lexer et al., 2003). The artificial development of the mapping populations, however, may generate genetic variation which may not occur in the wild (Erickson et al., 2004; Slate, 2005). Given the existence of several studies of individually-monitored, pedigreed individuals living in the wild, an obvious extension of these studies is to generate genetic maps and attempt to map genes underlying trait variation in nature. To date, however, we know of only two such studies pursuing this line, (excluding studies of humans, where cultural factors make extension of findings to animal populations difficult). In red deer (*Cervus elaphus*) living on the island of Rum SLATE *et al.* (2002) obtained a partial map (~62% genome coverage) using microsatellite marker genotypes, and then searched for QTL for a phenotypic trait, birth weight, finding three candidate regions for further investigation. Second, HANSSON *et al.* (2005) have recently generated a preliminary genetic map (~25% genome coverage) for the great reed warbler (*Acrocephalus arundinaceus*) population at Lake Kvismaren, Sweden, again using microsatellites.

In this paper, we describe the construction of a relatively much more complete genetic map for a free-living population, the Soay sheep (*Ovis aries*) living on St. Kilda,

taking advantage of existing genomic resources available for domestic sheep. This population is the subject of a long-term, individual-based multidisciplinary study which includes the collection of extensive phenotypic, ecological and genetic information (Clutton-Brock and Pemberton, 2004). Soay sheep are highly variable in appearance, with two independent polymorphisms of coat pigmentation (coat colour and coat pattern) and polymorphic horns (normal, deformed ("scurred") or polled horns). Selection acting on two of these polymorphisms, coat colour and horn type, has been previously documented (Milner et al., 2004). We demonstrate the utility of the genetic map by mapping the genes underlying these three polymorphic traits, setting the scene for better understanding of selection on these traits and for future QTL searches in the study population.

2.3. Materials and methods

Mapping population – The Soay sheep on the islands of Soay and Hirta (St. Kilda archipelago, North West Scotland, UK, 57°49' N, 08°34W) are feral populations of a breed regarded as the most primitive in Europe (Campbell, 1974; Doney et al., 1974); nowadays, the sheep population of Hirta varies between 600 and 2000 individuals. Since 1985 regular expeditions have been sent to St. Kilda to monitor the population dynamics and to record the entire history of individuals living in Village Bay, Hirta (Clutton-Brock et al., 2004a). No predators are present on St. Kilda.

The mapping population analyzed in this study was selected from a larger Soay sheep dataset comprising more than 3300 individuals with phenotypic records. In this population maternity is determined by observation, and paternity is inferred through molecular analysis (Overall et al., 2005). The mating system is polygynous and

promiscuous, such that very few full-sibs occur in the population. To trade off between power of linkage mapping and amount of genotyping work, we selected and genotyped half-sibships with twelve or more well-phenotyped individuals and their common parent, plus half-sibships with at least ten animals that were related to previously selected animals. In addition, we included in the mapping pedigree file, but did not genotype, the non-common parents and the ancestors of the half-sib families. Although not genotyped and in some cases phenotypically less well characterized, these additional individuals link different sibships which would appear otherwise as unrelated. This strategy increases the number of informative meioses as missing genotypes and marker phases can be, in some cases, inferred from the knowledge that different individuals share the same ancestors. In total, the mapping pedigree numbers 882 animals with 571 paternal links and 663 maternal links, of which 588 animals were genotyped (Figure 2.1).

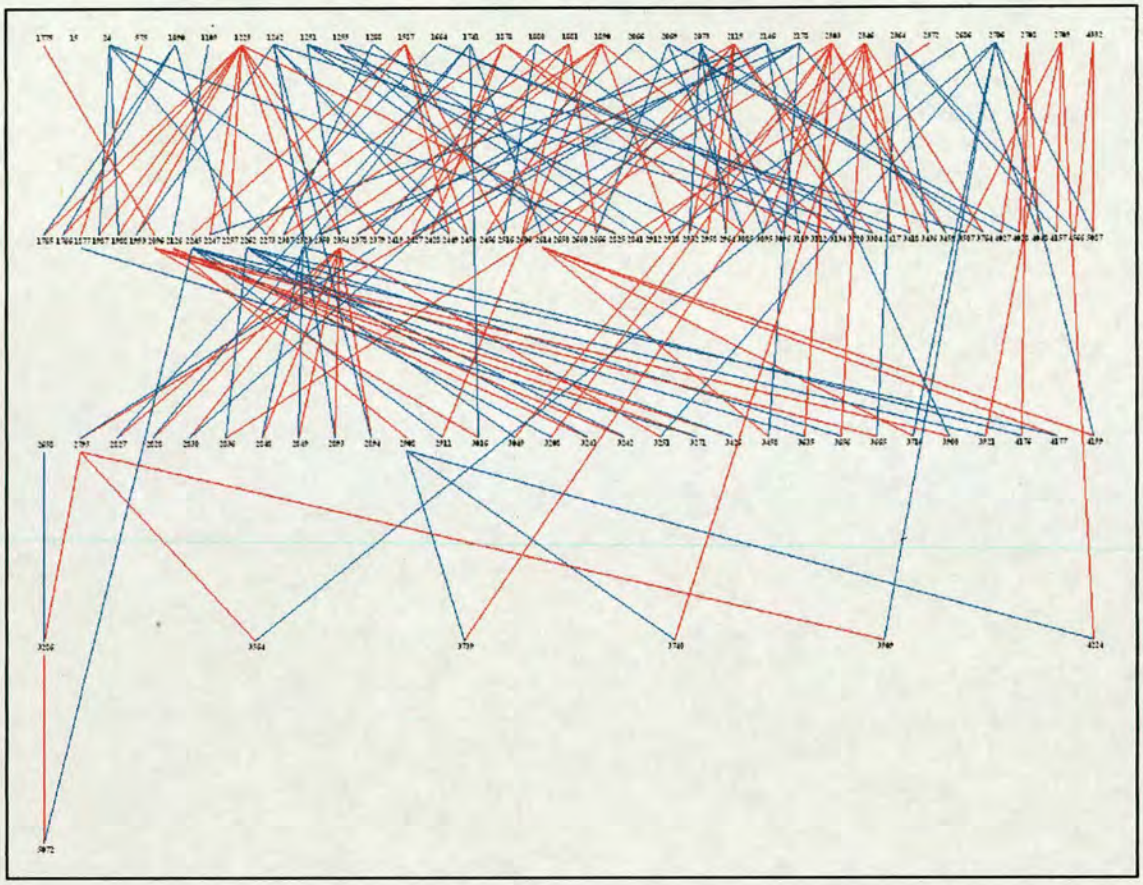

Figure 2.1 – Graphical representation of the mapping pedigree. 128 of the 588 genotyped members are displayed. 25 dams and 20 sires shown here have ≥ 10 offspring; the remaining individuals are additional dams and sires and individuals that link sibships. Blue lines: links from sires; red lines: links from dams. Numbers are the individual sheep ID numbers.

Polymorphic traits – In Soay sheep the colour of the pelage is determined by two independently segregating polymorphisms, one affecting the colour of the coat (hereafter referred to as coat colour, locus *Coat colour*, Figure 2.2A), the other determining the contrast in colour between belly and coat (hereafter referred to as coat pattern, locus *Coat pattern*, Figure 2.2B). Coat colour can be classified into two distinct phenotypes, dark and light which occur in a ratio of ~3:1. Segregation analyses in mainland Soays (Doney et al., 1974) and in resolved pedigrees on St. Kilda (Coltman and Pemberton, 2004) suggest that a single bi-allelic locus, in which dark is completely dominant to light, determines the two

classes (see Table 2.1). With respect to coat pattern, Soay sheep with the 'wild-type' morph have a paler belly and rump than the rest of the coat while the 'self' morph is characterised by a uniform and more intense coat colour. The wild and self morphs occur in a ratio of ~20:1. This variation is also determined by a single, bi-allelic locus with wild-type completely dominant to self (Coltman and Pemberton, 2004). Wild-type sheep have hairs in which the dark colour is alternated by pale bands, a pattern commonly found in wild mammals and usually due to the *Agouti* locus (Bennett and Lamoreux, 2003). Conversely, in self sheep the hairs have no banding pattern (Clutton-Brock et al., 2004b).

With respect to horn type (locus *Horn type*), Soay sheep are polymorphic for horns in both sexes. Females are classified into three horn types: normal (33%), scurred (vestigial and deformed, 28%) and polled (hornless, 39%); whereas in males only the normal (87%) and scurred (13%) phenotypes occur (Figure 2.2C, Table 2.1). Although the inheritance of the horn phenotype is not completely understood, pedigree data on St. Kilda is consistent with a single locus with three alleles (normal-horned, sex-limited horned and polled) showing sex-specific expression and dominance (Coltman and Pemberton, 2004), a model originally proposed for Merino sheep (Dolling, 1961).

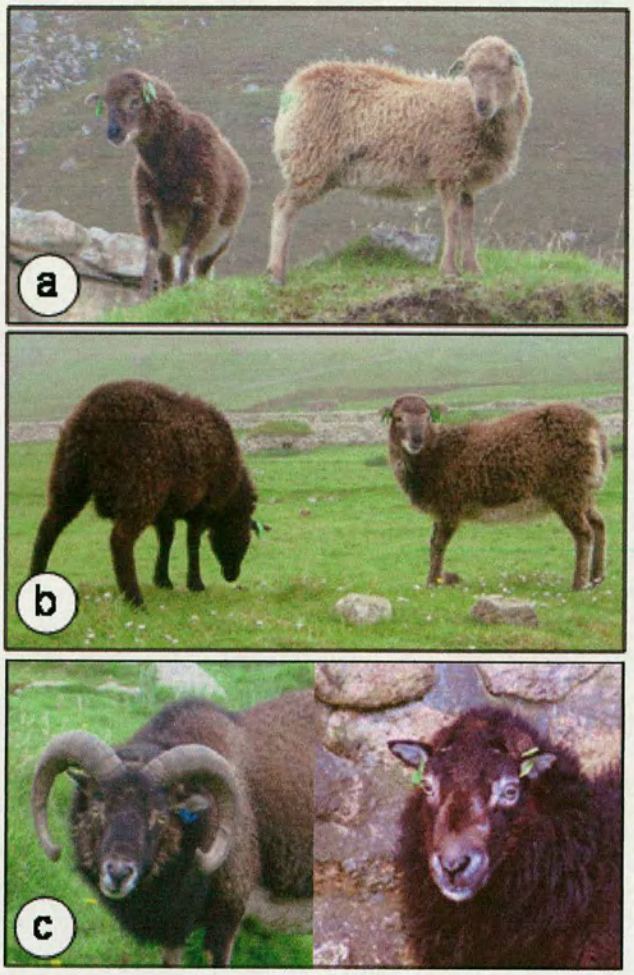

Figure 2.2 – Soay sheep showing the three traits subjected to linkage mapping. **A** – Coat colour polymorphism: dark (left) and light (right) lambs; **B** – Coat pattern polymorphism: self (left) and wild-type (right) lambs; note lack of contrast in colour between belly and rest of the body and the intensified coat colour in the self individual; **C** – Horn type morphs in adult males: normal (left) and extreme scurred (right).

Table 2.1. Phenotypic distributions and underlying genotypes of the study traits in the genotyped members of the Soay mapping pedigree (maximum n= 588).

Trait (n)	Phenotype	Genotype	Frequency
Coat colour (560)	Dark	*Dark*/-	0.74
	Light	*Light*/*Light*	0.26
Coat pattern (560)	Wild	*Wild*/-	0.94
	Self	*Self*/*Self*	0.06
Horn type – females (286)	Normal	Ho^+/Ho^+ Ho^+/Ho^L	0.38
	Scurred	Ho^L/Ho^L Ho^+/Ho^P	0.24
	Polled	Ho^P/Ho^P Ho^P/Ho^L	0.38
Horn type – males (270)	Normal	$Ho^+/-$ $Ho^L/-$	0.90
	Scurred	Ho^P/Ho^P	0.10

DNA extraction and microsatellite genotyping – Commercial kits were used to isolate DNA from blood samples (GFX Genomic Blood Purification Kit, Amersham Biosciences) or ear punches (GenomicPrep Cells and Tissue DNA isolation Kit, Amersham Biosciences) following the manufacturer's instructions. When the amount of starting material was too small or degraded to allow the use of these methods, the DNA was extracted using Chelex resin beads (Chelex 100 Resin, Bio-Rad Laboratories). About 1-5 mg of blood or tissue was incubated at 56°C overnight in 300 µl of a 5% Chelex and 0.1 µg/µl proteinase K solution followed by 5 min at 95°C. Before PCR amplification, the DNA solution extracted with either method was diluted 1:4 with ddH_2O and 2 µl was air-dried in 96 well PCR plates.

In order to construct a map with markers evenly distributed throughout the genome, the Australian Sheep Gene Mapping Web Site (Maddox, 2003) was taken as a reference to select microsatellite markers on the basis of their location and information content. PCR amplifications were performed in 5 μl volume, $MgCl_2$ concentration was adjusted between 1.5 and 4.0 mM to achieve optimal quality of the reaction. Two touch-down PCR programs were initially tested for each marker on a panel of 8 sheep: one in which the annealing temperature was progressively decreased during the first 10 of 29 cycles from 60 to 50°C, and the other in which the decrease was from 65 to 55°C. Fluorescent primers (6FAM, VIC, PET, or NED fluorescence) were synthesized by Applied Biosystems. Fragment lengths were analyzed on an ABI3730 DNA Analyzer and genotypes were determined using GeneMapper v3.0 (Applied Biosystems).

In order to estimate the genotyping error rate, 84 to 258 randomly chosen individuals were re-genotyped at 10 loci with average polymorphism. Genotyping error rate was also assessed on the basis of mother-offspring mismatches using CERVUS 2.0 (Marshall et al., 1998).

Linkage map and genome scan – Parent-offspring genotype inconsistencies arising from incorrect paternity assignment (32 incorrect links found) or typing errors were detected through the program PedCheck (O'Connell and Weeks, 1998) and either resolved by re-checking the parentage records and genotypes, or scored as untyped. Some cases of paternity mis-assignment were expected since in the original dataset paternity was assigned with only 80% confidence (Overall et al., 2005).

Linkage mapping was performed using CRI-MAP v2.4 (Green et al., 1990) to determine the marker order, inter-marker intervals, two-point LOD scores, and number of informative meioses. The complexity of the pedigree and the number of markers employed made impractical a systematic testing of all the possible map combinations. In

most cases the size of the pedigree did not allow the analysis of more than 7 or 8 markers at a time; therefore, sets of overlapping markers were tested sequentially until a whole chromosome was mapped. Markers expected to belong to the same chromosome were first input in CRI-MAP following the order reported in domestic sheep (Maddox, 2003). The log10-likelihood of the initial marker order was then compared with that of alternative orders (*flips2* or *flips3* function) to detect more likely combinations (i.e. higher log10-likelihood). An increase in log10-likelihood of three or more was considered as evidence of a significantly more probable map (Morton, 1955). In cases of inconsistency between Soay and domestic sheep, the most probable Soay order was retained after having ruled out possible human or technical mistakes. Markers mapping to unexpected locations or supported by a weak LOD score (<1.8) were also tested for linkage (*two-point* function) against all the other markers in the database to detect whether better positions could be found.

Coat colour and *Coat pattern* loci were mapped using CRI-MAP assuming each trait was encoded by a single locus with two alleles showing complete dominance: *Dark* allele dominant over *Light* and *Wild* dominant over *Self* (Coltman and Pemberton 2004, Table 2.1). A test for linkage between the target locus and any of the mapped markers was performed by means of the two-point function of CRI-MAP. The best position of a candidate locus was searched for by means of the *flips* and *fixed* functions of CRI-MAP to test alternative map positions. Consistent with the criteria used in the map construction, an increase in the log10-likelihood of the map of three or more was taken as evidence of a significantly more likely position. In the case of *Coat pattern*, the low frequency of *Self* morphs resulted in a low number of informative meioses. To confirm or reject a suggestive linkage, further sheep in families segregating for *Coat pattern* were genotyped at markers in the relevant region (see Results).

The *Horn type* locus was first investigated using CRI-MAP under a simplified model to scan the whole genome, and then the LINKAGE package (Terwilliger and Ott, 1994) was employed to perform parametric multipoint analysis in target region(s) identified by the preliminary scan. In CRI-MAP, *Horn type* was coded as a single bi-allelic locus where the *Normal* (Ho^+) allele was dominant and the *Polled* (Ho^P) allele was recessive in males (Ho^+ allele conferring normal horns when heterozygote or homozygote and Ho^P allele resulting in scurred horns when homozygote), whereas in females Ho^+ and Ho^P alleles were expressed co-dominantly (normal, scurred or polled horns given by Ho^+/Ho^+, Ho^+/Ho^P, Ho^P/Ho^P, respectively). This model was simplified in that the presumptive *Sex-limited* allele (Ho^L) was not explicitly considered (see Table 2.1); modelling three alleles in CRI-MAP would have resulted in too many missing genotypes since this program does not allow a trait (or a marker) phenotype to be coded by more than one genotype. Although this simplification reduces the power of analysis, it does not bias the results.

Computational constraints due to the size of the pedigree and the number of inbreeding loops prevented the processing of the entire pedigree by parametric multipoint linkage analysis. To circumvent this problem, the mapping panel was split into 39 unlinked families. For the parametric multipoint analysis in LINKAGE (Terwilliger and Ott, 1994), each sheep was assigned to one of five liability classes on the basis of their horn phenotype and sex: three classes for females (normal if Ho^+/Ho^+ or Ho^+/Ho^L, scurred if Ho^L/Ho^L or Ho^+/Ho^P, and polled if Ho^L/Ho^P or Ho^P/Ho^P; Table 2.1) and two classes for males (scurred horns if Ho^P/Ho^P, normal horns otherwise; males do not express the polled condition; Table 2.1), finally a sixth (fictitious) class was assigned to animals without phenotypic information; the underlying genotypes were assumed to have complete penetrance.

Horn allele frequencies were taken from Coltman and Pemberton (2004), as 0.441, 0.170 and 0.389 for Ho^{+}, Ho^{L} and Ho^{P}, respectively. Marker allele frequencies were estimated from the whole pedigree by 100,000 Markov chain Monte Carlo iterations implemented in Loki (Heath, 1997); this procedure is based on a stochastic process and as such does not provide an exact result, but allows the handling of very large and complex pedigrees.

The LOD threshold of 3.3 to declare evidence of linkage corresponds to the value usually applied to human pedigrees (Lander and Kruglyak, 1995). This decision was taken on the basis that the size of the sheep and human linkage maps are comparable.

2.4. Results

Soay sheep linkage map – The Soay sheep linkage map was developed with 247 microsatellite and 4 allozyme markers, giving a total of ~124,000 genotypes which generated a map with approximately 15 cM inter-marker spacing across 3350 cM, equivalent to ~3080 cM on the International Mapping Flock (IMF) map, and corresponding to ~90% of the sheep genome. Figure 2.3 compares the Soay sheep linkage map with the domestic sheep map (Maddox et al., 2001) and Table 2.2 in Appendix lists the mapped markers and their characteristics. The mean number of alleles per locus was 4.6 with a mean polymorphism information content (PIC) of 0.52 which are lower values than recorded on the Australian Sheep Gene Mapping Web Site for the same markers typed in the IMF (10 alleles and PIC=0.75); this is perhaps unsurprising since the latter figures are for a pedigree derived from several sheep breeds. On average, each marker was typed in 510 sheep (86% of the 588 sheep selected for genotyping) and generated 310 informative meioses. Genome wide, the mean LOD score for linkage

between two adjacent markers was 14.6. Twenty-two marker intervals were linked with a LOD score of less than 2, but their marker positions were retained since they were in agreement with the domestic sheep map (Maddox et al., 2001). Marker order was checked by means of the *flips* function of CRI-MAP and was consistent between the Soay and domestic sheep map in all but two cases: one on chromosome 1 where there is evidence for varying gene order in different sheep breeds (McRae and Beraldi, 2006), and the other on chromosome 12, which we have not investigated further. 2.4% of the 1290 duplicated genotypes showed inconsistency with the first screening. The error rate based on mother-offspring mismatching was 1.49% as estimated by CERVUS (Marshall et al., 1998).

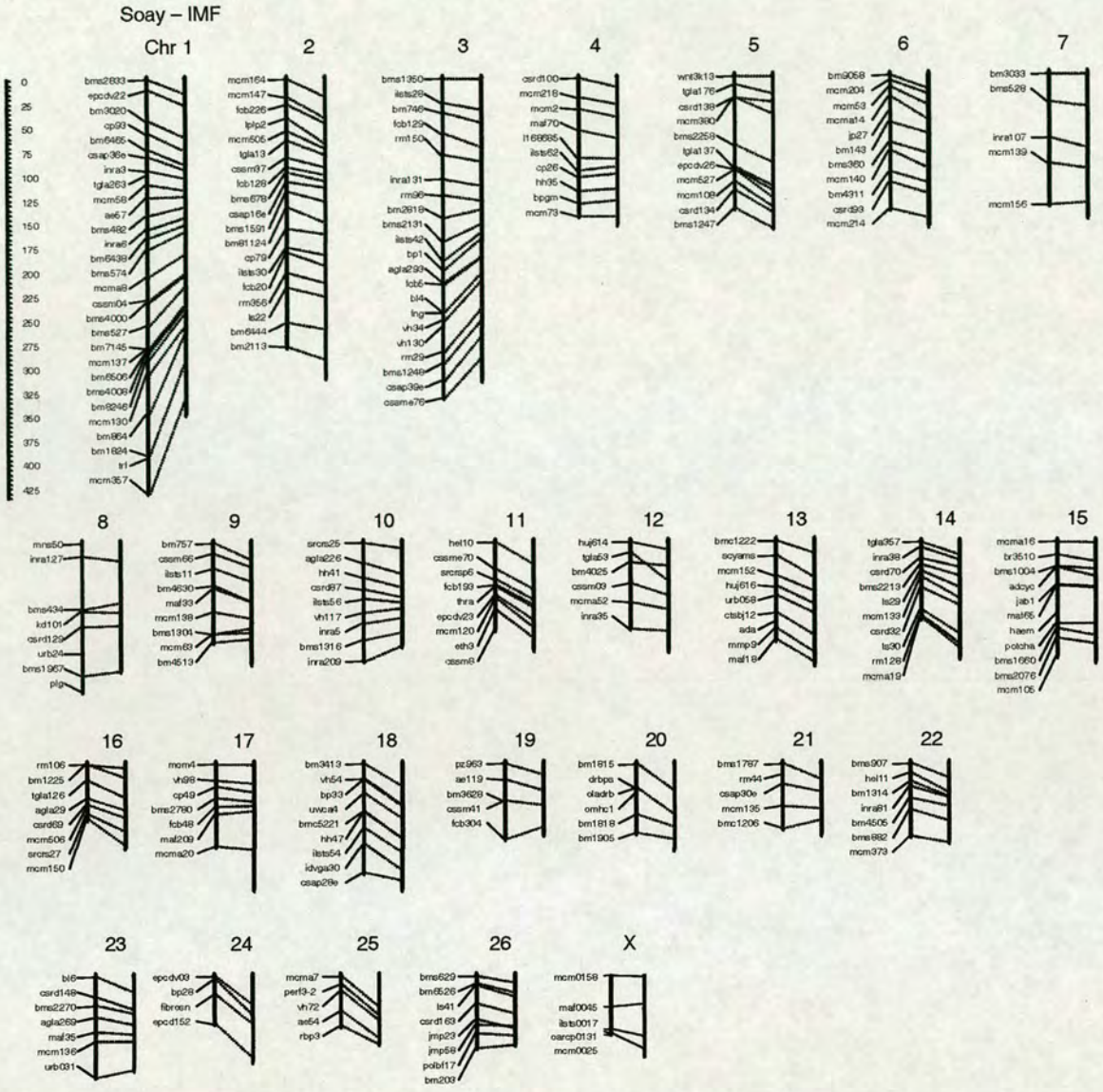

Figure 2.3 – Soay sheep map compared with the domestic (IMF) sheep map v4.3 (Maddox, 2003). In each pair, the Soay chromosome is on the left side, dotted lines connect markers shared by both maps. The ruler at the top left corner represents a centimorgan scale.

Linkage mapping of *Coat colour*, *Coat pattern*, and *Horn type* loci – The phenotypic distributions of *Coat colour*, *Coat pattern*, and *Horn type* in the mapping panel are reported in Table 2.1; these proportions do not differ significantly from the whole Soay sheep dataset (χ^2 test $p>0.1$).

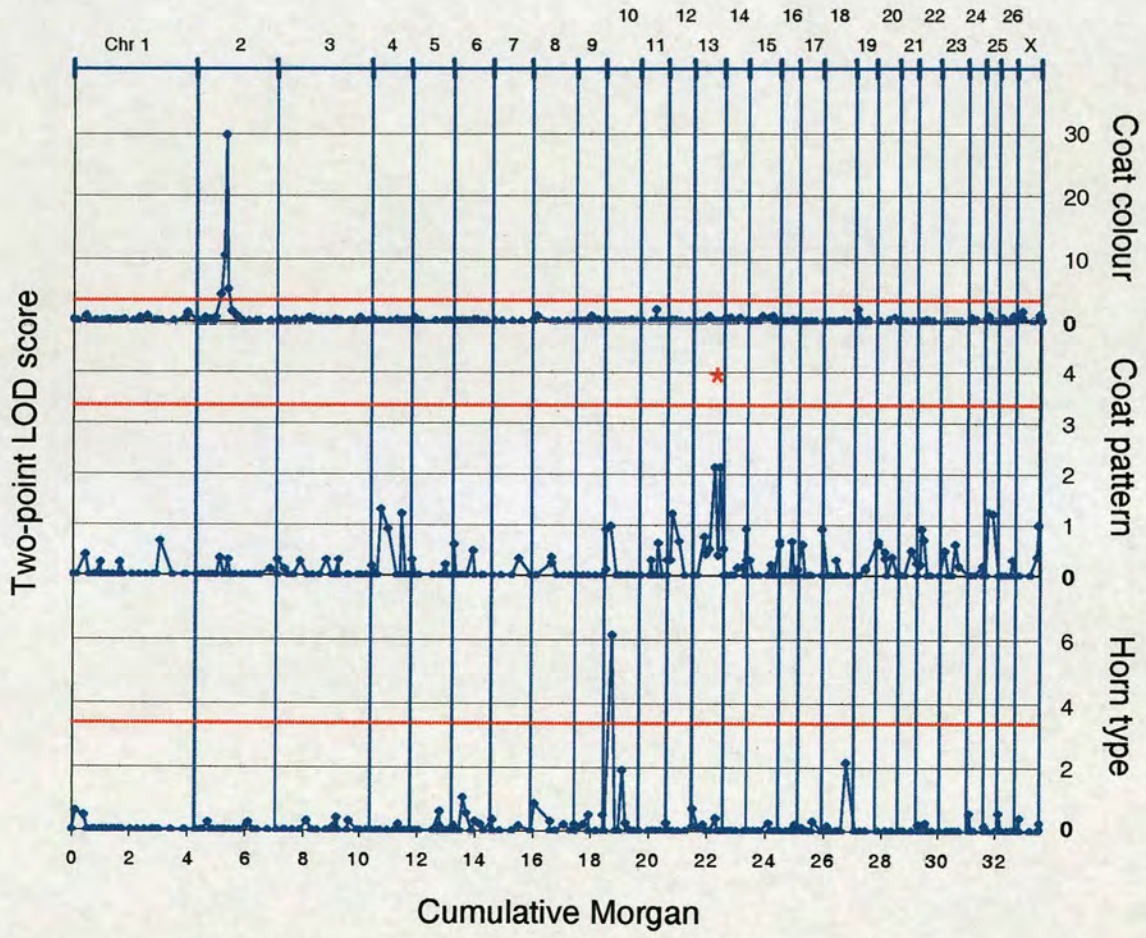

Figure 2.4 – Two-point LOD score profiles for linkage between study traits (reported on the right hand side of the graph) and markers (data points) in the Soay sheep map. The x axis represents the cumulative map distance of the genome (Morgan) with chromosome boundaries marked on top of the graph and by dotted vertical lines. Y axes report the LOD scale. The dotted lines denote the theoretical genome-wide significance threshold (LOD= 3.3). The asterisk in the middle panel shows the LOD score after having typed additional animals at marker CTSBJ12 (see text).

Coat colour: The highest LOD score for linkage was found with BMS678 (two-point LOD=29.5 at 0 cM), a microsatellite located on chromosome 2 (Figure 2.4). Other markers on chromosome 2 were significantly linked to the target locus, namely FCB128 (LOD=10.4), CSAP16E (LOD=5.1), and CSSM37 (LOD=4.4), whereas none of the other markers in the Soay sheep map produced a significant result for linkage (LOD<2). Figure 2.5A shows in detail the best position for the *Coat colour* locus in the map of chromosome 2, any other map order results in a significant decrease (>3) in the log10-likelihood of the map.

Coat pattern: The highest linkage score (LOD=2.1) was detected on chromosome 13 (Figure 2.4). This LOD score fell short of genome-wide significance, but this is likely to be a consequence of the low frequency of the self morph (6%), which meant that the *Coat pattern* locus was segregating in only a few families and there were few informative meioses for mapping (N=32). To confirm or reject this suggestive linkage, a further 78 animals comprising 15 families segregating for coat pattern were genotyped for the two microsatellite markers encompassing the LOD score peak (CTSBJ12 and MMP9), and the association between marker CTSBJ12 and the *Coat pattern* locus rose to LOD=3.9 with no recombinants between these loci (Figures 2.4 and 2.5B).

Horn type: Consistent with Montgomery et al. (1997), CRI-MAP detected linkage between *Horn type* and AGLA226 on chromosome 10 (LOD=6.1, Figure 2.4), but no other marker on chromosome 10 or elsewhere in the genome showed any significant linkage. Once the best location for *Horn type* was established on the chromosome 10 map (by use of the *fixed* function), CRI-MAP positioned *Horn type* distal to SRCRS25, the most telomeric marker on chromosome 10 (21.1 cM away from AGLA226). However, the likelihood of *Horn type* at this position was not significantly greater than in the interval between AGLA226 and SRCRS25 (log10-likelihood -150.16 versus -151.44), although

significantly better than in the interval AGLA226 – HH41 (log10-likelihood -153.17). Therefore, at this stage we concluded that *Horn type* is located on chromosome 10 distal to or in the vicinity of AGLA226 (Figure 2.5C), but an accurate map position could not be assigned.

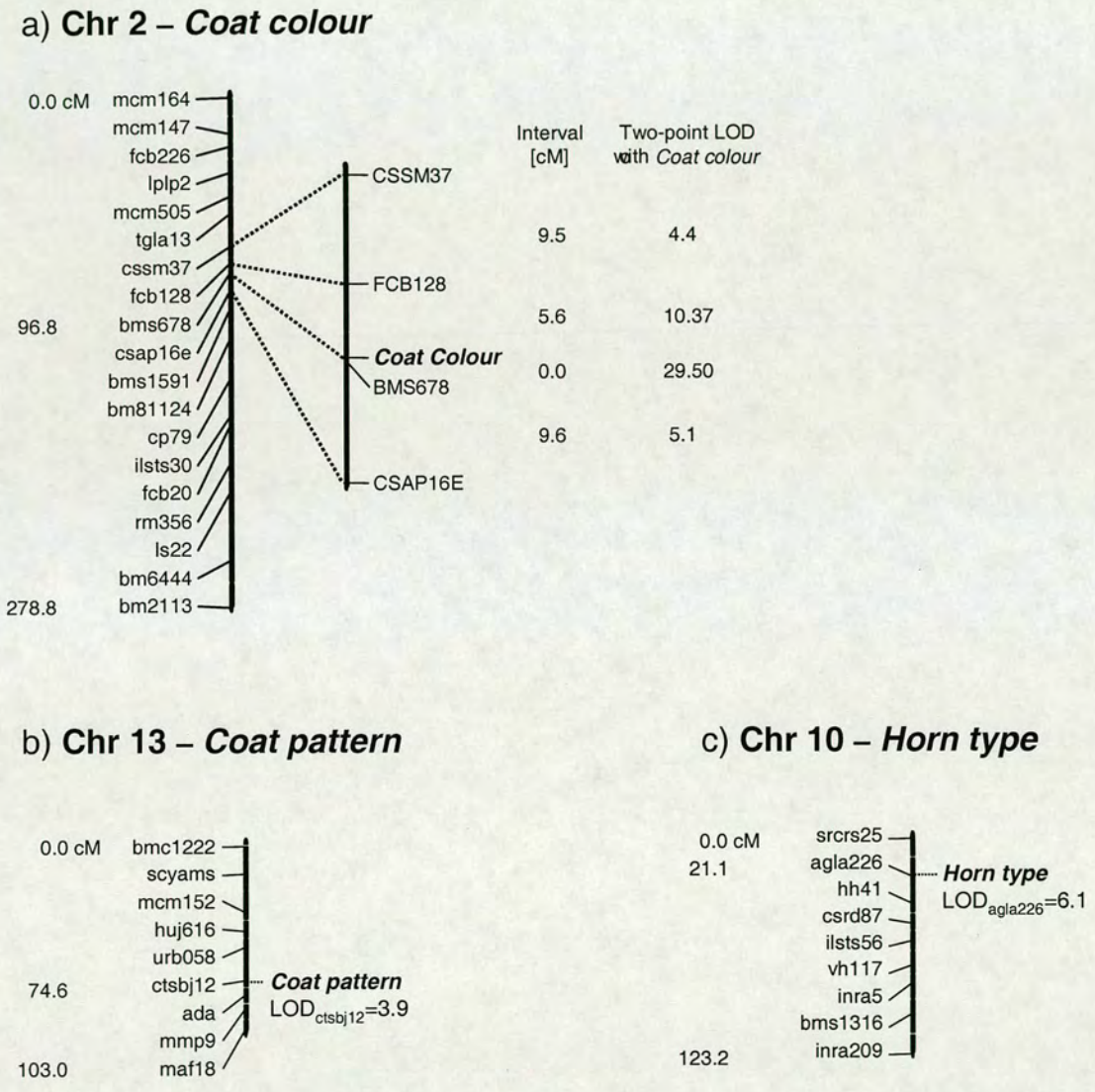

Figure 2.5 – Target regions identified by the genome scan for the three study traits. **A** – Chromosome 2 full map and detailed map of the region carrying the *Coat colour* locus; **B** – Suggestive region for the *Coat pattern* locus on chromosome 13; **C** – The *Horn type* location detected on chromosome 10 in the vicinity of AGLA226.

As described in the Methods, the CRI-MAP model of *Horn type* is simplified and does not account for the *Horn type* and marker allele frequencies. Therefore, the analysis was improved by performing multipoint parametric mapping to derive a more accurate estimate of the *Horn type* locus position. The *Horn type* locus was tested for linkage against AGLA226 and its two flanking markers. The LOD profile found by the multipoint analysis (Figure 2.6) suggests that the 1-LOD support interval for the presence of the target locus spans about 16 cM.

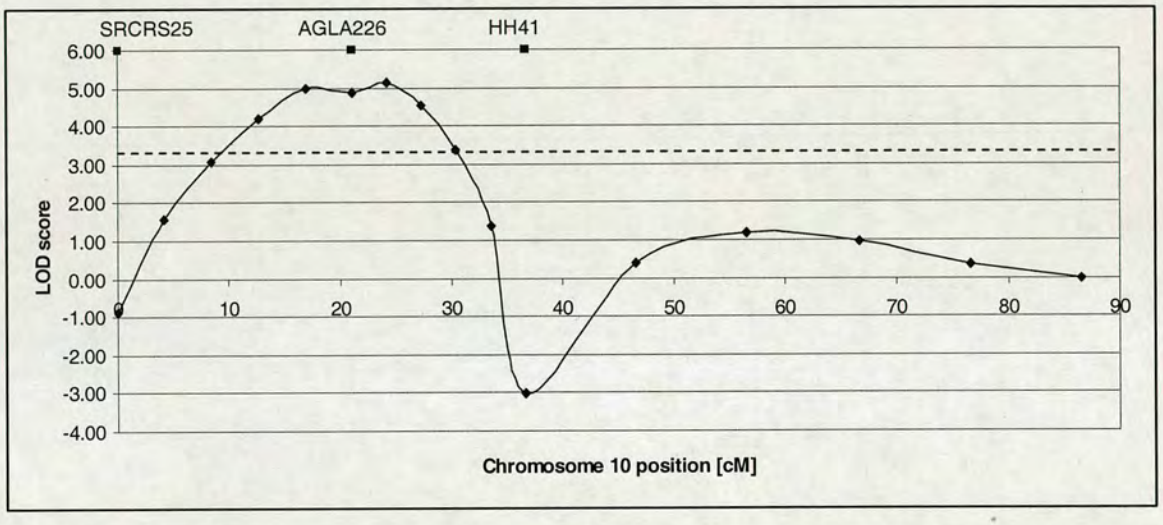

Figure 2.6 – Parametric four-point mapping of the *Horn type* locus. GLA226 on chromosome 10, the marker showing the strongest two-point linkage in CRI-MAP, and two adjacent markers (SRCRS25 and HH41) were simultaneously tested against *Horn type*. The location of the three markers is showed at the top. The *Horn type* position was tested every 5 cM (data-point). The dashed line denotes the theoretical genome-wide significance threshold (LOD= 3.3).

2.5. Discussion

As a step towards the comprehension of the genetic dynamics of wild populations, this article reports the development of a genetic map in a free-living population, the Soay sheep on St. Kilda, and its use in a genome scan to map the loci responsible for three morphological traits. To the best of our knowledge, this is one of the first

accomplishments of gene mapping in a free-living population. The Soay sheep on St. Kilda present interesting features from an evolutionary and genetic point of view: their number is naturally regulated by a combination of food availability, parasite burden, and winter weather (Coulson et al., 2001; Clutton-Brock et al., 2004a; Wilson et al., 2004), factors that, together, cause substantial fluctuations in population size (Coulson et al., 2001; Clutton-Brock et al., 2004a).

Development of the Soay sheep linkage map – The map presented here has been developed with the primary purpose of localizing genes of evolutionary interest. Especially when the phenotype conveys little information about the underlying genotype, as is the case for many quantitative characters, the monitoring of the target trait is improved and complemented by the genotype inferred through linked markers.

Patterns of allelic association in terms of linkage disequilibrium and population structure provide insights into history and selection of a population (Abecasis et al., 2005). To this end, a linkage map is a starting point to enrich regions of interest with markers in order to assess the extension of the association and to compare this latter with theoretical expectations (McRae et al., 2005). Genomic tools such as comparative mapping will facilitate the discovery of additional markers and candidate genes in target regions.

Mapping of *Coat colour, Coat pattern, Horn type* – The attempt to map the locus responsible for coat colour variation successfully yielded a region on chromosome 2 (Figure 2.4) defined by a window of approximately 15 cM (Figure 2.5A) in which the *Coat colour* locus co-segregates with BMS678. Gratten et al. (In press) following a candidate gene method have tested for association with different genes known to affect

coat colour in mammals, and they identified the responsible gene (TYRP1) and its causal mutation.

The interest for coat colour in Soay sheep stems from the differential survival between dark and light animals although no predators are present on St. Kilda and no obvious environmental conditions should favour one colour over the other. It has previously been found that dark coats are positively selected during some high mortality winters, but this is inconsistent and in other winters selection favours light coloured sheep or neither morph (Moorcroft et al., 1996; Milner et al., 2004). Dark animals are significantly heavier than light ones, providing a possible mechanism for their better survival (Clutton-Brock et al., 1997). There is no difference in female fecundity between dark and light sheep (Clutton-Brock et al., 1997). At present, there is no explanation for why the light colour morph is maintained in the population; clearly, being able to distinguish the three genotypes may shed light on this puzzle. Hypotheses and future work to explain the difference in survival will take advantage of the map position and molecular characterization of the *Coat colour* gene. A comparison between LD in the FCB128 – CSAP16E interval and background LD in the Soay sheep genome should also provide information about the origin and evolutionary consequences of coat colour variation.

With respect to *Coat pattern*, the high frequency of the wild morph (94% of the sheep scored, Table 2.1) severely reduced the number of informative meioses (32) so that strong linkage to any marker was unlikely to be found. The power of linkage mapping is proportional to the fraction of parents heterozygous at both the target locus and at the linked markers. This combination generates the necessary marker-trait association in the progeny (Lynch and Walsh, 1998). It follows that if the target locus has a highly skewed allelic distribution, few heterozygous individuals are generated and more meioses need to

be scored (the information content, estimated as PIC, reaches the highest value when all the alleles have the same frequency). Accordingly, the highest LOD score for *Coat pattern* reached only 2.1 on chromosome 13 (Figure 2.4) after an initial scan. However, the extension of the sample size confirmed this suggestive linkage. Interestingly, chromosome 13 harbours the *Agouti* locus, a candidate for *Coat pattern* (Parsons et al., 1999). *Agouti* encodes for an antagonist of the melanocortin receptor causing a switch from eumelanin to pheomelanin production in the pigment producing cells, which results in the characteristic banding pattern observed in Soay sheep hairs and other mammals (Bennett and Lamoreux, 2003). To date we have not detected selection acting on the *Coat pattern* locus.

Multipoint parametric linkage analysis was not performed for *Coat colour* and *Coat pattern* because, in contrast to *Horn type*, the CRI-MAP model for *Coat colour* and *Coat pattern* was already consistent with the most likely model, so that little or no improvement would have been gained by multipoint parametric analysis.

The mapping of *Horn type* returned a telomeric region on chromosome 10 previously detected by Montgomery et al. (1996, Figure 2.5C and 2.6). The present work opens the way for multiple strategies to fine map and isolate the *Horn type* gene. These include: exploitation of bioinformatics tools to enrich the target region with SNPs and other microsatellites, and identification of positional candidates by comparison with the annotated genome assemblies of cattle and other species. Like coat colour, horn phenotype is under selection in Soay sheep and other wild populations. In ruminants, horns are typically used in intrasexual conflict, particularly among males where they reach much greater size. Previous analyses of Soay sheep have suggested that normal-horned males and scurred females have highest annual breeding success (Clutton-Brock et al., 1997; Stevenson et al., 2004), but that in winters characterized by high mortality, the

scurred phenotype is generally favoured in both sexes (Moorcroft et al., 1996). Exactly how these forces maintain variation in the population is the subject of current research and would clearly be helped by being able to distinguish individuals by genotype rather than phenotype. Therefore, the *Horn type* region is an attractive target for molecular evolution studies.

Future directions – The traits analyzed here are characterized by relatively simple inheritance patterns which, to some extent, may limit their applicability to the understanding of the process of evolution. However, this project opens the way to the more challenging task of detecting QTL affecting a variety of morphological and physiological traits. The Soay sheep has been the subject of a number of studies aimed at estimating quantitative genetic parameters for traits like birth weight and body size (Coltman et al., 1999; Milner et al., 2000; Milner et al., 2004). It has been found that the additive genetic variance of these traits is low but not null, despite the pressure of selection acting on them (Milner et al., 2000). As these previous studies have been conducted under the infinitesimal model framework, the dissection of these traits through QTL mapping to determine eventual Mendelian factors would represent a major breakthrough towards the comprehension of the evolutionary processes in the wild.

Acknowledgements

We thank J.G. Pilkington and many volunteers for collecting field data and genetic samples. We thank the National Trust for Scotland for granting permission to work on St. Kilda, and QinetiQ for logistical support. We thank two anonymous reviewers for their comments on the first version of this article. The long term data collection on St. Kilda has been supported by Natural Environment Research Council (NERC) and

Wellcome Trust grants to T.H. Clutton-Brock, B.T. Grenfell, L.E.B. Kruuk, M.J. Crawley and J.M. Pemberton. This study was funded by NERC through its Environmental Genomics thematic programme (grant number NER/T/S/2002/00189).

2.6. Appendix

Table 2.2. Details of the markers included in the Soay sheep genetic map

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
1	BMS2833	40.43	11	0	533	496	5	0.762	0.743	0.7	0.0053	14.2	9	0.7
1	EPCDV22	10.17	32	11	543	463	6	0.773	0.743	0.707	0	26.4	7	NA
1	BM3020	24.99	13.6	43	481	440	6	0.778	0.785	0.75	0.0052	59.5	13	0.74
1	CP93	16.03	15.7	56.6	545	343	4	0.618	0.593	0.55	0	73.6	6	0.58
1	BM6465	42.6	8.2	72.3	522	405	6	0.707	0.737	0.697	0.0107	87.7	8	NA
1	CSAP36E	10.8	13.8	80.5	525	484	5	0.73	0.739	0.692	0.0058	93.5	10	0.82
1	INRA3	20.68	4.3	94.3	517	216	2	0.453	0.455	0.352	0	102.8	5	0.46
1	TGLA263	30.03	10	98.6	402	377	7	0.816	0.787	0.755	0	NA	NA	NA
1	MCM58	20.1	13.8	108.6	550	464	7	0.78	0.746	0.702	0	114	14	0.88
1	AE57	3.48	18.9	122.4	502	290	4	0.552	0.534	0.474	0.0545	120.7	9	0.76
1	BMS482	12.81	13.6	141.3	528	281	4	0.424	0.457	0.429	0.0707	132.3	11	NA
1	INRA6	12.53	8.7	154.9	500	267	3	0.582	0.563	0.466	0	142.7	8	0.52
1	BM6438	9.71	12.9	163.6	526	302	4	0.551	0.535	0.454	0	150.6	6	0.77
1	BMS574	4	28.6	176.5	434	238	4	0.562	0.581	0.518	0.0164	157.5	12	0.75
1	MCMA8	5.03	23.5	205.1	544	215	3	0.406	0.435	0.394	0	181.3	11	0.65
1	CSSM04	17.51	2.9	228.6	528	236	3	0.492	0.452	0.359	0	200.9	8	NA
1	BMS4000	3.47	23.6	231.5	425	202	5	0.595	0.569	0.496	0	204.1	15	0.82
1	BMS527	7.6	24.2	255.1	526	292	5	0.555	0.535	0.491	0.0117	214.9	12	0.82
1	BM7145	89.4	0.8	279.3	503	394	4	0.718	0.721	0.669	0	234.6	6	0.66
1	MCM137	69.12	2.1	280.1	558	523	7	0.763	0.752	0.715	0	233.4	15	0.85
1	BM6506	58.31	1.4	282.2	506	333	5	0.686	0.676	0.63	0.0075	235.6	6	NA

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
1	BMS4008	35.13	6.8	283.6	522	411	7	0.787	0.765	0.728	0	236.7	10	0.75
1	BM8246	10.2	15.6	290.4	525	350	5	0.617	0.599	0.558	0.009	242.2	9	NA
1	MCM130	0	40.7	306	553	363	4	0.571	0.541	0.435	0.0114	256.5	16	0.73
1	BM864	0	44.3	346.7	520	143	3	0.287	0.284	0.256	0	264.5	11	0.73
1	BM1824	40.75	3.1	391	501	322	5	0.653	0.671	0.621	0	294.1	5	0.68
1	TRF	0	36.8	394.1	539	437	7	0.785	0.764	0.728	0.0135	NA	NA	NA
1	MCM357	0		430.9	517	176	2	0.422	0.417	0.33	0.0631	332.6	11	0.85
2	MCM164	3.74	19.2	0	546	97	3	0.192	0.181	0.169	0	17.6	13	0.78
2	MCM147	8.58	6.8	19.2	533	414	7	0.717	0.722	0.68	0.0114	39.8	12	0.83
2	FCB226	3.49	14.7	26	529	127	4	0.386	0.377	0.348	0	44.6	12	0.79
2	LPLP2	4.35	13	40.7	546	441	6	0.749	0.739	0.696	0.0051	65.5	14	0.84
2	MCM505	3.15	10.2	53.7	551	80	5	0.181	0.182	0.176	0	71.4	8	0.7
2	TGLA13	8.47	17.8	63.9	397	277	4	0.718	0.737	0.687	0.0165	77.8	6	NA
2	CSSM37	30.66	9.5	81.7	545	374	5	0.556	0.566	0.537	0	91.1	12	0.54
2	FCB128	48.92	5.6	91.2	544	439	4	0.662	0.645	0.597	0.0241	99.4	8	0.72
2	BMS678	16.38	9.6	96.8	549	426	5	0.619	0.614	0.542	0	106.1	14	0.83
2	CSAP16E	6.24	9.8	106.4	523	146	2	0.283	0.285	0.244	0.0468	112.4	6	0.43
2	BMS1591	10.86	17.1	116.2	484	264	5	0.556	0.521	0.481	0.0131	127	16	0.85
2	BM81124	9.89	21.8	133.3	526	410	4	0.679	0.638	0.566	0	148.2	11	NA
2	CP79	4.61	19.8	155.1	473	306	6	0.668	0.667	0.611	0.0248	161	8	NA
2	ILSTS30	8.73	5.4	174.9	550	120	2	0.213	0.21	0.188	0	182.4	10	0.7
2	FCB20	7.47	20.6	180.3	400	236	7	0.61	0.635	0.6	0.075	194	12	0.8
2	RM356	13.78	15.6	200.9	496	360	6	0.752	0.748	0.705	0.0244	211.1	16	NA
2	LS22	3.41	36.2	216.5	546	349	5	0.668	0.638	0.57	0	225.8	8	0.78

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
2	BM6444	7.72	26.1	252.7	521	425	6	0.758	0.755	0.713	0	260.2	13	0.88
2	BM2113			278.8	495	356	4	0.614	0.595	0.514	0.0218	291.9	8	0.67
3	BMS1350	4.76	25	0	438	187	5	0.393	0.403	0.377	0	0	12	NA
3	ILSTS28	15.04	9.9	25	553	356	5	0.633	0.607	0.56	0	32	14	0.82
3	BM746	5.17	21.9	34.9	537	287	3	0.529	0.537	0.43	0.0243	45.4	9	0.52
3	FCB129	3.49	22.7	56.8	528	347	4	0.595	0.625	0.577	0.0084	70.2	10	0.71
3	RM150	2.3	24.8	79.5	536	160	5	0.326	0.34	0.316	0	85.9	6	NA
3	INRA131	12.56	16.5	104.3	523	327	3	0.639	0.657	0.582	0.0278	111.4	7	NA
3	RM96	3.81	23.4	120.8	521	344	4	0.601	0.55	0.494	0.0129	126.7	6	0.56
3	BM2818	2.89	24.7	144.2	496	214	3	0.528	0.531	0.45	0.0726	134.9	5	NA
3	BMS2131	6.94	19.8	168.9	543	327	3	0.652	0.634	0.559	0.0753	150.4	5	NA
3	ILSTS42	15.04	9.9	188.7	530	365	7	0.675	0.654	0.612	0.007	159.1	10	NA
3	BP1	4.84	12.6	198.6	521	257	2	0.53	0.5	0.375	0	166	4	0.38
3	AGLA293	10.34	2.7	211.2	557	241	2	0.434	0.433	0.339	0.0525	184.7	4	0.36
3	FCB5	0.61	28.6	213.9	547	163	2	0.333	0.313	0.264	0	186.3	3	0.42
3	BL4	36.8	1	242.5	569	301	3	0.559	0.581	0.488	0.0095	202.5	8	NA
3	FNG	12.43	7.4	243.5	563	238	2	0.448	0.465	0.357	0	NA	NA	NA
3	VH34	0	30.3	250.9	572	329	5	0.579	0.576	0.532	0	210.7	9	0.68
3	RM29	0.75	11.7	281.2	177	69	5	0.644	0.592	0.53	0.0347	238.6	8	0.67
3	BMS1248	5.83	17.9	292.9	530	204	2	0.413	0.421	0.332	0.0588	251.1	11	0.67
3	CSAP39E	2.41	17.4	310.8	517	236	3	0.596	0.545	0.482	0.0761	268.7	11	0.75
3	CSSME76			328.2	542	106	3	0.231	0.236	0.22	0	290.4	9	0.68
4	CSRD100	9.49	19.5	0	549	207	3	0.364	0.364	0.309	0	9.2	8	0.71
4	MCM218	38.21	12	19.5	529	437	5	0.773	0.748	0.705	0.0052	26.5	9	0.82

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
4	MCM2	10.53	22.5	31.5	555	502	7	0.791	0.796	0.771	0.0038	39.4	11	0.8
4	MAF70	1.7	28.6	54	381	263	6	0.743	0.785	0.749	0.1455	61.4	17	0.9
4	L168685	7.76	12.9	82.6	462	220	4	0.545	0.559	0.503	0.0268	83.3	9	NA
4	ILSTS62	28.98	7	95.5	517	332	5	0.617	0.602	0.543	0	91.3	16	0.88
4	CP26	18.59	14.4	102.5	560	422	5	0.738	0.723	0.674	0.0054	98.7	6	0.74
4	HH35	21.57	13.3	116.9	548	408	4	0.695	0.696	0.643	0.0125	114.9	7	0.7
4	BPGM	15.12	12.5	130.2	545	394	6	0.653	0.641	0.611	0	126.1	7	0.7
4	MCM73			142.7	545	316	3	0.486	0.489	0.402	0.0851	143.6	13	0.88
5	WNT3K13	14.56	9.3	0	550	191	4	0.438	0.415	0.386	0	0	8	0.76
5	TGLA176	9.37	12.3	9.3	549	292	3	0.574	0.538	0.464	0	17.8	8	NA
5	MCM380	9.63	0	21.6	559	70	4	0.114	0.116	0.111	0	26.6	10	0.78
5	CSRD138	0	50	21.6	535	383	5	0.6	0.577	0.519	0	38.7	10	0.76
5	BMS2258	5	22.7	71.6	551	266	2	0.454	0.479	0.364	0	89.6	8	NA
5	TGLA137	26.63	2.2	94.3	516	288	3	0.523	0.517	0.462	0	113.2	8	NA
5	EPCDV26	32.74	1.9	96.5	553	302	5	0.582	0.58	0.542	0	117.4	7	NA
5	MCM527	27.7	12.1	98.4	397	324	6	0.783	0.758	0.727	0.0152	125.5	6	0.67
5	MCM108	30.28	8.8	110.5	545	503	7	0.811	0.784	0.753	0	135.2	11	0.82
5	CSRD134	5.36	18.8	119.3	501	281	4	0.559	0.613	0.557	0.0286	140.8	6	0.55
5	BMS1247			138.1	545	370	3	0.662	0.629	0.552	0.0086	157	7	NA
6	BM9058	22.59	5	0	452	312	5	0.615	0.624	0.578	0	12.9	10	NA
6	MCM204	20.7	6.8	5	541	318	4	0.54	0.516	0.465	0	18.2	8	0.79
6	MCM53	6.29	10.7	11.8	558	378	5	0.67	0.664	0.623	0.0064	29.7	9	0.76
6	MCMA14	2.09	14.5	22.5	543	88	3	0.179	0.178	0.17	0	45	8	0.67
6	JP27	25.46	11.9	37	526	449	8	0.793	0.798	0.768	0.0084	NA	NA	NA

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
6	BM143	10.29	20.7	48.9	544	332	6	0.563	0.549	0.506	0	59	10	NA
6	BMS360	20.64	8.3	69.6	536	388	4	0.688	0.671	0.609	0	80.8	12	NA
6	MCM140	5.28	22.3	77.9	447	317	5	0.609	0.599	0.533	0.0123	95.8	10	0.79
6	BM4311	10.78	10.7	100.2	447	355	5	0.996	0.705	0.647	0	111.6	9	0.79
6	CSRD93	5.77	27.3	110.9	537	427	4	0.695	0.7	0.646	0	122.7	11	0.73
6	MCM214			138.2	552	417	4	0.679	0.658	0.597	0	147.1	7	0.72
7	BM3033	5.02	28.2	0	541	379	3	0.632	0.634	0,562	0.0084	0	17	NA
7	BMS528	1.7	38.2	28.2	533	348	3	0.612	0.618	0.536	0.0097	32.6	6	NA
7	INRA107	3.82	26.6	66.4	517	386	3	0.598	0.574	0.49	0.0358	75.4	17	0.79
7	MCM139	1.2	42.6	93	518	236	5	0.542	0.555	0.524	0.0203	97.6	11	0.81
7	MCM156			135.6	534	294	4	0.614	0.632	0.572	0.0756	133.4	8	0.77
8	MNS50A	10.11	13.3	0	547	463	5	0.779	0.737	0.688	0.0165	5	11	0.85
8	INRA127	0	54.6	13.3	503	211	3	0.342	0.366	0.301	0	17.2	8	0.81
8	BMS434	62.99	2.4	67.9	515	394	6	0.724	0.692	0.652	0.0069	61.3	9	NA
8	KD101	16.03	16.8	70.3	552	452	6	0.766	0.758	0.716	0.0048	71.1	12	0.83
8	CSRD129	2.99	27.2	87.1	541	328	5	0.636	0.627	0.566	0	86	11	0.82
8	URB024	9.86	23.5	114.3	528	270	4	0.532	0.549	0.49	0	117.5	12	0.65
8	BMS1967	6.09	16.1	137.8	535	503	6	0.776	0.769	0.731	0	132.8	11	0.71
8	PLG			153.9	277	144	2	0.596	0.5	0.375	0	NA	NA	NA
9	BM757	26.84	10.9	0	550	364	4	0.675	0.67	0.615	0.0069	16.2	7	NA
9	CSSM66	12.42	15.4	10.9	550	325	4	0.578	0.569	0.507	0.0309	24.2	10	NA
9	ILSTS11	9.6	18.2	26.3	533	290	3	0.465	0.449	0.401	0	40.1	10	0.68
9	BM4630	6.18	2.6	44.5	540	368	6	0.548	0.611	0.555	0.0451	60	8	0.78
9	MAF33	0	24.4	47.1	193	36	4	0.285	0.302	0.286	0.2015	60	9	0.7

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
9	MCM138	1.12	23.1	71.5	539	222	2	0.341	0.418	0.33	0.1234	78.5	9	0.64
9	BMS1304	3.31	0.4	94.6	45	33	6	1	0.782	0.737	0	89.6	12	NA
9	MCM63	29.71	8	95	546	362	5	0.632	0.645	0.602	0.016	93.1	14	0.87
9	BM4513			103	429	403	7	0.811	0.806	0.776	0	100.3	10	0.82
10	SRCRS25	7.22	21.1	0	553	253	4	0.483	0.498	0.463	0.0515	5.3	11	0.82
10	AGLA226	10.64	15.5	21.1	481	312	6	0.636	0.68	0.646	0.0078	31.5	8	0.78
10	HH41	9.83	11.7	36.6	545	366	7	0.615	0.634	0.585	0	45.9	11	0.8
10	CSRD87	14.02	9.8	48.3	538	291	4	0.61	0.585	0.504	0	56.2	10	0.78
10	ILSTS56	22.31	14.7	58.1	500	359	6	0.682	0.72	0.687	0.0621	61.2	9	0.7
10	VH117	19.16	10.7	72.8	541	451	6	0.726	0.734	0.693	0.0108	68.2	11	0.75
10	INRA5	11.47	17.8	83.5	403	314	8	0.695	0.679	0.635	0.0181	78	13	0.87
10	BMS1316	5.93	21.9	101.3	547	452	5	0.682	0.673	0.609	0	93.5	13	0.81
10	INRA209			123.2	546	179	2	0.264	0.297	0.253	0.043	108	10	0.58
11	HEL10	4.2	27.7	0	550	349	5	0.645	0.624	0.588	0.0077	22.4	12	0.85
11	CSSME70	4.33	10.7	27.7	550	261	4	0.58	0.574	0.498	0	47.9	9	0.78
11	SRCRSP6	10.56	5.8	38.4	537	394	4	0.68	0.663	0.612	0	59.9	9	0.73
11	FCB193	7.63	2.3	44.2	564	146	3	0.229	0.226	0.203	0	65.4	12	0.41
11	THRA	12.8	8.9	46.5	545	258	3	0.499	0.497	0.387	0.0138	67.2	8	0.73
11	EPCDV23	61.23	3	55.4	546	371	7	0.689	0.665	0.624	0	79.9	10	NA
11	MCM120	27.89	12.2	58.4	550	490	8	0.773	0.76	0.725	0	87.7	16	0.85
11	ETH3	3.57	20.6	70.6	536	319	3	0.565	0.557	0.491	0	99.4	5	NA
11	CSSM08			91.2	541	229	4	0.447	0.43	0.404	0	112.4	5	0.53
12	HUJ614	1.99	10.5	0	543	65	2	0.145	0.141	0.131	0	7.6	9	0.72
12	TGLA53	15.23	9.9	10.5	400	308	7	0.64	0.651	0.621	0.039	39.3	8	NA

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
12	BM4025	5.29	22.4	20.4	445	319	5	0.712	0.714	0.659	0.0074	24	9	NA
12	CSSM03	12.42	19.8	42.8	535	229	3	0.393	0.396	0.36	0	54.5	10	NA
12	MCMA52	3.5	27.4	62.6	536	387	5	0.674	0.678	0.624	0.0069	68.8	10	0.78
12	INRA35			90	523	212	5	0.512	0.477	0.436	0	92.1	9	0.68
13	BMC1222	19.35	15.5	0	508	416	5	0.734	0.706	0.67	0	12.3	15	0.82
13	SCYAMS	14.79	20.4	15.5	535	407	5	0.693	0.683	0.642	0.0197	37.4	20	0.89
13	MCM152	23.42	10.9	35.9	527	356	4	0.651	0.635	0.56	0.0087	52.1	10	0.79
13	HUJ616	13.51	10.3	46.8	546	365	5	0.615	0.609	0.538	0.0092	65	15	NA
13	URB058	4.59	17.5	57.1	548	254	4	0.5	0.508	0.387	0.0261	74.4	13	0.78
13	CTSBJ12	5.63	8.9	74.6	552	154	4	0.252	0.249	0.222	0	98	9	0.77
13	ADA	11.85	8.1	83.5	456	217	2	0.382	0.38	0.307	0	NA	NA	NA
13	MMP9	8.94	11.4	91.6	542	359	5	0.646	0.635	0.585	0	115.4	9	0.79
13	MAF18			103	335	171	3	0.573	0.559	0.488	0.019	125.8	5	0.41
14	TGLA357	41.95	5.1	0	525	315	6	0.651	0.657	0.606	0	11.7	8	0.8
14	INRA38	15.8	11	5.1	536	398	6	0.668	0.69	0.643	0	17.6	13	0.82
14	CSRD70	25.28	7.2	16.1	530	358	5	0.675	0.656	0.608	0	25.5	14	0.78
14	BMS2213	27.13	5.9	23.3	530	325	4	0.587	0.59	0.52	0	33.8	10	0.82
14	LS29	17.05	7.2	29.2	536	367	6	0.631	0.614	0.575	0	46.5	14	0.84
14	MCM133	11.53	11.2	36.4	532	185	3	0.361	0.362	0.31	0.0272	56.8	9	0.75
14	CSRD32	8.34	23	47.6	527	293	4	0.545	0.511	0.459	0	64.6	15	0.82
14	LS30	21.14	6.5	70.6	532	308	4	0.609	0.635	0.568	0	94.4	11	0.79
14	RM128	36.86	2.2	77.1	556	303	4	0.482	0.473	0.436	0	104.6	11	0.81
14	MCMA19			79.3	519	189	3	0.374	0.354	0.326	0	109.3	6	NA
15	MCMA16	15.86	12.9	0	551	372	4	0.633	0.631	0.575	0.0081	0	9	0.63

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
15	BR3510	3.04	12.9	12.9	515	289	5	0.561	0.542	0.502	0	19.3	8	NA
15	BMS1004	6.33	0.98	25.8	532	99	2	0.195	0.189	0.171	0	27.2	13	NA
15	ADCYC	21.26	17	26.78	539	379	5	0.651	0.65	0.585	0	35.3	7	0.74
15	JAB1	23.14	1.2	43.78	533	388	8	0.683	0.703	0.655	0.0061	46.5	19	0.82
15	MAF65	0	29.8	44.98	328	172	4	0.518	0.512	0.453	0	47	8	0.62
15	HAEM	2.29	9.8	74.78	216	57	2	0.528	0.501	0.375	0	NA	NA	NA
15	POTCHA	30.64	5.3	84.58	510	413	5	0.716	0.717	0.668	0	85.1	11	NA
15	BMS1660	13.4	10	89.88	532	282	4	0.515	0.484	0.442	0	96.5	10	0.77
15	BMS2076	4.04	19.1	99.88	531	301	3	0.605	0.606	0.537	0	105.4	14	0.83
15	MCM105			118.98	534	500	8	0.785	0.784	0.75	0.0044	123.8	10	0.81
16	RM106	12.64	0	0	327	130	4	0.474	0.48	0.447	0	3.8	10	NA
16	BM1225	0	16.5	0	543	133	3	0.274	0.267	0.249	0	13.2	9	0.74
16	TGLA126	6.98	18.5	16.5	525	212	4	0.371	0.383	0.342	0.0247	34.3	18	0.84
16	AGLA29	30.04	6.7	35	502	308	6	0.631	0.637	0.581	0.0083	46.9	16	NA
16	CSRD69	6.28	3.7	41.7	519	361	6	0.667	0.681	0.636	0.0073	55.3	10	0.67
16	MCM506A	4.85	6.6	45.4	533	73	3	0.126	0.12	0.116	0	63.2	17	0.75
16	SRCRS27	4.58	5.2	52	509	385	6	0.68	0.688	0.631	0	69.4	8	0.72
16	MCM150			57.2	544	42	2	0.131	0.147	0.136	0	83.9	9	0.68
17	MCM4	0	16.9	0	550	404	5	0.596	0.662	0.616	0	0	13	0.85
17	VH98	9.19	6	16.9	553	71	4	0.179	0.178	0.172	0	19.7	9	0.67
17	CP49	22.83	12.6	22.9	543	449	7	0.703	0.687	0.649	0.0063	28.5	7	0.76
17	BMS2780	3.23	8.8	35.5	517	289	4	0.509	0.516	0.439	0.0147	38.4	8	0.75
17	FCB48	7.12	7.4	44.3	302	109	4	0.45	0.457	0.405	0.1119	42.5	11	0.76
17	MAF209	5.31	34	51.7	402	332	7	0.756	0.738	0.692	0.0515	48	8	0.79

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
17	MCMA20			85.7	507	489	6	0.984	0.806	0.777	0	89.2	11	0.73
18	BM3413	6.73	15.5	0	555	237	4	0.373	0.378	0.327	0	22.4	8	0.7
18	VH54	49.16	1.5	15.5	554	320	4	0.605	0.625	0.555	0	41.9	7	0.7
18	BP33	25.55	13.9	17	549	438	6	0.719	0.715	0.675	0	43.4	12	0.85
18	UWCA4	14.29	18.5	30.9	551	401	4	0.641	0.634	0.584	0	56.1	6	0.72
18	BMC5221	31.87	1.4	49.4	551	409	5	0.713	0.715	0.666	0	77	10	NA
18	HH47	2.5	16	50.8	379	205	5	0.623	0.658	0.607	0.0108	77	10	0.77
18	ILSTS54	0	16.8	66.8	552	89	2	0.239	0.248	0.217	0.1071	91.7	3	NA
18	IDVGA30	0	29.5	83.6	533	150	2	0.257	0.344	0.285	0.3782	110.5	2	NA
18	CSAP28E			113.1	558	261	4	0.461	0.439	0.393	0.0169	121.6	5	0.62
19	PZ963	12.91	16.9	0	549	350	6	0.65	0.618	0.574	0	10.6	22	NA
19	AE119	3.98	21.6	16.9	523	298	3	0.535	0.513	0.459	0	27.7	8	0.76
19	CSSM41	9.33	0	38.5	551	151	2	0.385	0.39	0.314	0.0219	NA	NA	NA
19	BM3628	2.47	40.1	38.5	499	264	5	0.515	0.511	0.47	0	43.3	4	NA
19	FCB304			78.6	577	352	4	0.591	0.586	0.506	0	66	9	0.54
20	BM1815	6.78	24.8	0	544	255	3	0.504	0.521	0.405	0.0867	26.8	6	NA
20	OLADRB	48.16	0	24.8	529	478	8	0.79	0.819	0.794	0.0149	52.2	13	NA
20	OLADRBps	23.85	5.6	24.8	263	203	6	0.795	0.786	0.752	0.014	NA	NA	NA
20	OMHC1	4.58	22	30.4	294	222	5	0.568	0.597	0.564	0.0325	NA	NA	NA
20	BM1818	6.45	19.2	52.4	533	388	8	0.657	0.678	0.629	0.0848	64.9	10	NA
20	BM1905			71.6	562	320	2	0.528	0.488	0.369	0	77.8	2	NA
21	BMS1787	7	11	0	524	438	4	0.716	0.711	0.656	0	15.5	16	0.84
21	RM044	5.07	9.8	11	520	125	3	0.26	0.251	0.23	0	22.1	10	0.83
21	CSAP30E	4.62	23.1	20.8	540	269	3	0.48	0.508	0.401	0	29.1	15	0.79

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
21	MCM135	2.84	22.9	43.9	493	399	5	0.996	0.676	0.618	0.0093	46	13	0.85
21	BMC1206			66.8	538	212	2	0.522	0.499	0.374	0	58.1	6	0.67
22	BMS907	48.34	10.1	0	498	480	7	0.815	0.838	0.817	0.0083	13.8	12	0.83
22	HEL11	24.14	9.1	10.1	531	509	8	0.825	0.849	0.83	0.0154	30	17	NA
22	BM1314	34.63	4.6	19.2	333	288	8	0.802	0.81	0.784	0	34.5	6	NA
22	INRA81	29.94	12.9	23.8	504	420	7	0.734	0.772	0.733	0	35.7	22	0.85
22	BM4505	32.57	11.9	36.7	518	402	7	0.707	0.734	0.69	0.0058	43.5	11	NA
22	BMS882	5.18	27.4	48.6	545	485	5	0.734	0.734	0.691	0	59.7	7	0.79
22	MCM373			76	531	231	6	0.405	0.411	0.393	0.021	82.9	13	0.84
23	BL6	8.13	20.8	0	520	310	7	0.619	0.636	0.576	0.0091	15.7	12	NA
23	CSRD148	12.28	9.2	20.8	502	319	5	0.697	0.673	0.616	0	33.2	15	0.82
23	BMS2270	9.05	11.7	30	557	224	5	0.361	0.361	0.339	0	37.7	8	0.79
23	AGLA269	13.22	15.2	41.7	508	333	5	0.646	0.621	0.561	0	49.1	18	NA
23	MAF35	10.49	10.5	56.9	583	369	3	0.568	0.565	0.473	0	59.2	5	0.61
23	MCM136	0	37.6	67.4	552	438	5	0.737	0.743	0.694	0	67.6	8	0.83
23	URB031			105	516	148	4	0.347	0.369	0.347	0.0567	97	7	0.62
24	EPCDV03	45.75	3.6	0	529	344	4	0.59	0.565	0.508	0.0108	28.9	8	NA
24	BP28	21.85	14.8	3.6	540	400	4	0.67	0.664	0.611	0.0143	38.5	18	0.92
24	FIBROSN	1.9	31.4	18.4	549	421	4	0.727	0.724	0.674	0.0057	48.9	12	0.82
24	EPCD152			49.8	525	439	5	0.985	0.656	0.587	0	83.2	6	NA
25	MCMA7	12.21	8.4	0	528	224	3	0.422	0.391	0.337	0	31	12	0.9
25	PERF3-2	16.39	6.8	8.4	525	382	5	0.714	0.671	0.615	0.0077	40	15	0.86
25	VH72	3.18	21.1	15.2	522	171	4	0.276	0.26	0.247	0	45.6	7	0.76
25	AE54	3.97	15.1	36.3	402	231	5	0.617	0.632	0.567	0.0442	64	9	0.82

												IMF Map[12]		
Chr[1]	Marker	2ptLOD[3]	Inter[2]	Pos[4]	N[5]	InfMei[6]	No All[7]	H(O)[8]	H(E)[9]	PIC[10]	Est Err Rate[11]	Pos[4]	No All[7]	PIC[10]
25	RBP3			51.4	332	200	3	0.642	0.608	0.536	0	69.9	4	0.61
26	BMS629	19.53	8	0	527	309	3	0.575	0.579	0.49	0	6.9	9	NA
26	BM6526	26.85	2.3	8	541	336	4	0.534	0.536	0.48	0.0121	16.4	9	NA
26	LS41	7.1	18.1	10.3	535	310	4	0.624	0.617	0.542	0.0091	20.6	12	0.77
26	CSRD163	17.18	16.9	28.4	524	270	3	0.578	0.614	0.531	0.0097	39	8	NA
26	JMP23	20.7	5.3	45.3	531	443	8	0.727	0.731	0.681	0.0058	53.5	13	NA
26	JMP58	9.71	8.9	50.6	401	222	5	0.589	0.596	0.522	0.0263	51.4	9	0.67
26	POLBF17	11.2	14.8	59.5	497	328	7	0.678	0.669	0.641	0	61.5	11	0.82
26	BM203			74.3	335	278	9	0.773	0.79	0.759	0.0424	71.1	10	NA
X	MCM158	4.27	34.6	0	528	397	7	0.714	0.648	0.599	0	0.8	12	0.88
X	MAF45	18.21	24.7	34.6	709	623	6	0.779	0.742	0.701	0	31.2	12	0.84
X	ILSTS17	66.49	5.1	59.3	530	543	3	0.577	0.47	0.41	0	66.8	11	0.79
X	CP131	72.24	1.8	64.4	524	537	7	0.517	0.525	0.494	0	80.5	9	0.83
X	MCM25			66.2	528	464	5	0.729	0.628	0.579	0	90.8	15	NA
	Average	14.58	15.0		510	310	4.58	0.58	0.58	0.52	0.01		10.06	0.75
	St Dev	14.88	9.9		73	113	1.58	0.17	0.16	0.16	0.03		3.49	0.11

[1] Chromosome number; [2] Inter-marker spacing (cM); [3] LOD score for linkage between adjacent markers; [4] Chromosomal position (cM); [5] Number of sheep genotyped; [6] Number of informative meioses detected at the marker locus; [7] Number of alleles; [8] Observed marker heterozygosity; [9] Estimated marker heterozygosity; [10] Polymorphism information content; [11] Estimated error rate from mother-offspring pairs using CERVUS; [12] IMF map characteristics (Maddox 2003).

Chapter 3.

Mapping QTL underlying fitness-related traits in a free-living sheep population

Dario Beraldi[1,4], Allan F. McRae[1,2,5], Jacob Gratten[3,6], Jon Slate[3,7], Peter M. Visscher[1,2,8] and Josephine M. Pemberton[1,9]

[1] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, UK.*

[2] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, 4029, Australia.*

[3] *Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK.*

[4] dario.beraldi@ed.ac.uk; [5] allan.mcrae@qimr.edu.au; [6] j.gratten@sheffield.ac.uk; [7] j.slate@sheffield.ac.uk; [8] peter.visscher@qimr.edu.au; [9] j.pemberton@ed.ac.uk

Author contribution: DB performed the statistical analysis, and wrote the manuscript. AFM and PMV helped with the statistical analysis. All the contributing authors critically read and improved drafts of this manuscript. The whole project was planned and supervised by JMP, JS and PMV.

3.1. Abstract

We searched for QTL underlying fitness related traits in a free-living pedigree of 588 Soay sheep in which a genetic map using 251 markers with average spacing 15 cM had been established previously. Traits examined included birth date and weight, considered both as maternal and offspring traits, foreleg length, hindleg length and body weight measured on living or immediately post mortem individuals and jaw length and metacarpal length measured on cleaned skeletal material. In some cases the data were split to consider different age classes separately, yielding a total of 15 traits studied. Genetic and environmental components of phenotypic variance were estimated for each trait and, for those traits showing non-zero heritability (N = 12), a QTL search was conducted by comparing a polygenic model with a model including a putative QTL. Support for a QTL at genome-wide significance was found on chromosome 11 for jaw length; suggestive QTL were found on chromosomes 2 (for birth date as a trait of the lamb), 8 (birth weight as a trait of the lamb) and 15 (adult hindleg length). We discuss the prospects for refining estimates of QTL position and effect size in the study population, and for QTL searches in free-living pedigrees in general.

Keywords: variance components, QTL, birth date, birth weight, body size, Soay sheep.

3.2. Introduction

Quantitative variation arising from the combined effects of many genes and the environment is a common feature of all populations of eukaryotes and the subject of intensive genetic research across many fields. From an evolutionary perspective, heritable variation is essential for the progress of evolution by natural selection, as only the genetic material can be transferred from one generation to another (Fisher, 1958).

There has been extensive debate on the amount of heritable variation expected in quantitative traits in natural populations. Since natural selection is often directional, i.e. it favours the individuals at one end of the phenotypic distribution of a trait (Endler, 1986), it is expected that selection depletes disadvantageous alleles and leads to the fixation of favourable variants. Eventually, this process should remove the genetic diversity in a population, and the reduction of variability should be faster for traits more closely related to fitness. A survey of quantitative genetics parameters from a number of free-living, outbred animal populations has confirmed that the more closely related a trait is to fitness, the lower its heritability (Mousseau and Roff, 1987; Roff and Mousseau, 1987; Roff and Simons, 1997). However, in contrast to predictions, if the additive genetic variances of different traits are standardized by the trait mean rather than by the total variation, it appears that fitness-related traits tend to have *higher* additive genetic and non-additive genetic components than traits less closely related to fitness (Houle, 1992; Kruuk et al., 2000; McCleery et al., 2004). This counterintuitive finding suggests that fitness traits have a broad biological and genetic basis which confers high additive genetic variation, and it is the high environmental variation in fitness-related traits which result in their having relative low heritability (Houle, 1992).

The study of quantitative traits is traditionally based on a theoretical framework, the infinitesimal model, where no knowledge of the number and location of the genes that underlie them is required (Fisher, 1918). With respect to gene distribution and action across the genome, it is sufficient to assume that many genes are segregating in a given population, and that their individual allelic differences are small relative to the effects of the environment. Although the infinitesimal model is applicable to a wide variety of studies including natural populations, it fails to capture the complexity of genetic variation in terms of the number of genes involved, their relative effect, and their action and interaction. As early as 1973, for example, Elston and Stewart (Elston and Stewart, 1973; Stewart and Elston, 1973) realized that a continuous normal phenotypic distribution is not necessarily generated by many additive loci but it can result from the action of a small number of genes interacting with the environment. Determining whether or not there are detectable QTL segregating within natural populations, and if detected estimating their effect sizes, is thus of central interest to the study of the evolution of quantitative traits by natural selection.

The development of techniques to generate and screen large numbers of molecular markers in many individuals has opened the way to investigating the whole genome in order to test whether specific regions (Quantitative Trait Loci, QTL) affect variation in a trait more than the average background. The most insightful successes in QTL mapping have been recorded in experimental populations (Paterson et al., 1991; Kroymann and Mitchell-Olds, 2005). A typical experiment aimed at mapping QTL starts with crossing inbred lines which have been selected to diverge for the trait of interest; consequently, the selected lines are expected to be homozygous for the alleles conferring the extreme phenotypes. The mapping individuals, often F_2 or backcross progenies, are then raised under controlled and uniform conditions designed to enhance the expression of the

target phenotype and to reduce environmental heterogeneity (Lynch and Walsh, 1998). Since the probability of QTL detection increases with the heritability of the study trait (Williams and Blangero, 1999), such experimental designs offer high statistical power to map QTL: in the segregating progeny the additive genetic variance, i.e. the detectable genetic factor, is maximized while the environmental variance is reduced.

Unfortunately, the findings discovered with these approaches cannot be readily used to interpret natural variation in the wild because the experimental conditions are often over-simplified or unrealistic. An experimental design like the one described above has extremely reduced genetic diversity since only the alleles carried by the parental lines are screened in the segregating progeny. Furthermore, artificial crosses, such as inter-specific crosses, may generate variation (i.e. segregating QTL) at loci which do not segregate within a population in the wild. In addition, interactions between genotype and environment are probably altered. Although different strategies have been devised to overcome these limitations, such as the crossing and growing of wild individuals under controlled conditions (e.g. Lexer et al., 2003), the direct study of natural populations in their own environment would be the least dependent on compromising assumptions and constraints.

Mapping QTL in wild populations poses several practical and analytical difficulties (Slate, 2005). Any linkage mapping project requires samples from related individuals in order to generate a linkage map, and the same individuals should have phenotypic records of interest available with, ideally, environmental sources of variation known. The gathering of this information is far from trivial in a wild population, especially with respect to the reconstruction of accurate pedigrees. Statistical analysis may also become computationally demanding when data present an unbalanced structure or a high level of

complexity. In this respect, the statistical procedures designed by animal or plant breeders are not optimal for natural populations as they usually require balanced data and are performed on single sibships or simple pedigrees (e.g. least square regression, Haley et al., 1994). For these reasons few projects have been undertaken to dissect natural genetic variation through QTL mapping (for an exception see Slate et al., 2002) despite the recognized interest (Erickson et al., 2004). The human population can be considered 'wild' and under natural selection in many respects such as disease resistance (Olson, 2002; Sabeti et al., 2002; Bamshad and Wooding, 2003), and linkage mapping in humans has proved to be a useful strategy to understand the genetic basis of complex traits, especially medical conditions (Abiola et al., 2003; Botstein and Risch, 2003). However, the peculiar cultural background of the human population and the use of medical treatments make extension of findings to non-human populations difficult.

The free-living Soay sheep population on Hirta, St. Kilda, UK, is the subject of a long term project aimed at addressing a wide range of ecological and evolutionary issues (extensively documented in Clutton-Brock and Pemberton 2004). The population dynamics are characterized by periodic fluctuations in the number of individuals. The population size increases until the density of animals exceeds the winter carrying capacity and, as a result, a large proportion of individuals die in the following winter due mainly to starvation. High mortality is exacerbated by wet and windy winters and heavier burdens of gastrointestinal parasitic nematodes. Mortality is age- and sex-related in that juveniles are more susceptible to poor environmental conditions and males experience lower over-winter survival, probably due to energy expenditure during the Autumn rut (Coulson et al., 2001; Clutton-Brock et al., 2004a). The variable conditions of early growth and

survival set up substantial differences in many traits between sheep cohorts (Clutton-Brock and Pemberton, 2004)

The Soay sheep population expansions and contractions offer opportunities for natural selection. Previous studies have investigated the genetic basis of a variety of traits associated with fitness in Soay sheep (Clutton-Brock and Pemberton, 2004). Traits measured in early development, i.e. birth date and birth weight, contribute to total fitness in Soay sheep (Clutton-Brock et al., 1992; Jones et al., 2005) and in other mammals such as red deer (Kruuk et al., 1999). In Soay sheep, lambs born heavier and later in the season have better survival in the first few months of life (Wilson et al., 2005b). Early development continues to influence survival during the first winter, when early-born and below average weight neonates are more likely to die, especially if the population density is high (Clutton-Brock et al., 1992; Milner et al., 1999). A detailed analysis of selection acting via mothers and offspring suggests there are differences in trait optima for mothers and offspring (Wilson et al., 2005b). Beyond the neonatal period, body size is one of the most important characteristics of an organism because many other morphological, physiological, behavioural and life history traits scale with it (Schmid et al., 2002). In common with other mammal populations such as bighorn sheep (Coltman et al., 2002) and red deer (Kruuk et al., 1999), fitness of Soay sheep is related to body size measured either as body weight or hindleg length (a measure of skeletal size). In winters characterized by high mortality, directional selection favours heavier and longer-legged individuals (Milner et al., 1999). Heavier females are more fecund and more successful at rearing offspring (Clutton-Brock et al., 1997). Larger males (in terms of both body weight and hindleg length) have higher reproductive success, through an increased ability to monopolize receptive females (Preston et al., 2003).

Previous quantitative genetic analyses in the Soay sheep population indicate that the fitness-related traits discussed above harbour additive genetic variation. Heritability estimates for birth date and weight are 0.06 and 0.08 respectively, but in each case there is also a relatively substantial maternal genetic component (0.28 and 0.12 respectively as a proportion of phenotypic variance) (Wilson et al., 2005a). These two traits also show substantial genetic correlation ($R_G = 0.962 \pm 0.375$ (SE)). In two studies using different data subsets, estimates for the heritability of body weight at age 4 months and above were 0.24 and 0.28 for females and 0.12 and 0.07 for males (Milner et al., 2000; Coltman et al., 2001a) and heritability estimates for hindleg length were 0.26 and 0.35 for females and 0.20 and 0.24 in males (Milner et al., 2000; Coltman et al., 2001a). These two traits again show substantial genetic correlation (R_G estimates for females 0.74 ± 0.09 and 0.80 ± 0.02 and for males 0.78 ± 0.10 and 0.78 ± 0.05)

Altogether, the large volume phenotypic, pedigree and environmental data that has been collected since 1985 makes the Soay sheep suitable for linkage mapping projects. Elsewhere we have described the construction of a genetic map for the study population and the mapping of three Mendelian traits segregating in Soay sheep (Beraldi et al., 2006). Here we report the result of variance components analyses and genome scans aimed at identifying QTL for neonatal traits (birth weight, birth date), and body size (body weight, hind and fore leg length, jaw length and metacarpal length). The results presented constitute one of the first attempts to dissect the complexity of quantitative traits in the wild, and the methods here applied, based on an extension of methods developed for complex traits in humans (Amos, 1994; Almasy and Blangero, 1998; Williams and Blangero, 1999; George et al., 2000), should be suitable for the analysis of any natural population with a pedigree of similar or even higher complexity.

3.3. Materials and methods

Study Population – The Soay sheep on the islands of Soay and Hirta (St. Kilda archipelago, North West Scotland, UK, 57°49' N, 08°34W) are feral populations of a breed regarded as the most primitive in Europe (Campbell, 1974; Doney et al., 1974); nowadays, the sheep population of Hirta varies between 600 and 2000 individuals. Since 1985 regular expeditions have been sent to St. Kilda to monitor the population dynamics and to record the life histories of individuals living in Village Bay, Hirta (Clutton-Brock and Pemberton, 2004). No predators are present on St. Kilda.

Mapping Pedigree and Linkage Map – The whole Soay sheep pedigree file numbers more than 3900 animals. Within this pedigree maternal links were assigned through observation of the animals in the field, whereas paternal links were inferred through molecular analysis (Overall et al., 2005). From the total pedigree, a panel of 588 animals was genotyped at 247 microsatellite and four isoenzyme markers. This subset comprised all the sibships with ten or more offspring and their common parents. The ancestors of the genotyped individuals and the animals linking different sibships (n= 294) were not genotyped, but they were included in the mapping pedigree to improve the estimates of the kinship and identity by descent (IBD) coefficients in the variance components analysis. A more thorough description of the mapping pedigree and selection criteria is included in Beraldi et al. (2006). The Soay sheep map covers approximately 90% of the genome with an average inter-marker spacing of 15 cM. Further details of the map characteristics and of the technical procedures can be found in Beraldi et al. (2006).

Phenotypic Dataset – Phenotypic records of the animals in the mapping pedigree were retrieved from the Soay sheep database where data for more than 6000 sheep are

stored. The data analysed in this study were collected between 1988 and 2005 from animals born between 1978 and 2002. Sample sizes and summary statistics for each trait are reported in Table 3.1.

Table 3.1. Characteristics and estimated variance components of the study traits (NS: non significant)

Trait	Dataset	No. records	No. animals (genotyped)	Mean (SE[a])	V_A[b] (SE[a])	CV_A[c] (%)	h^{2d} (SE[a])	V_M[e] (SE[a])	CV_M[f] (%)	m^{2g} (SE[a])	V_C[h] (SE[a])	CV_C[i] (%)	c^{2j} (SE[a])	V_E[k] (SE[a])	CV_E[l] (%)	e^{2m} (SE[a])
Birth date (lamb)	Neonatal	710	710 (526)	111 (8)	6.467 (4.042)	-	*0.07 (0.04)*	63.946 (7.944)	-	*0.69 (0.04)*	NS	-	*NS*	22.434 (3.299)	-	*0.24 (0.04)*
Birth date [days] (mother)	Neonatal	1901	311 (136)	140 (8)	19.700 (6.274)	-	*0.28 (0.08)*	NS	-	*NS*	10.776 (5.107)	-	*0.16 (0.07)*	38.756 (0.393)	-	*0.56 (0.03)*
Birth weight [kg] (lamb)	Neonatal	601	601 (507)	2.30 (0.61)	0.022 (0.012)	6.44	*0.16 (0.09)*	0.034 (0.009)	8.06	*0.25 (0.06)*	NS	NS	*NS*	0.079 (0.010)	12.26	*0.58 (0.08)*
Birth weight [kg] (mother)	Neonatal	1708	306 (133)	2.22 (0.73)	0.048 (0.007)	9.86	*0.27 (0.03)*	NS	NS	*NS*	NS	NS	*NS*	0.131 (0.005)	16.34	*0.73 (0.03)*
Foreleg length [mm]	All	2242	737 (509)	124 (10)	6.977 (2.491)	2.14	*0.16 (0.06)*	1.116 (0.908)	0.86	*0.03 (0.02)*	18.334 (2.423)	3.47	*0.44 (0.06)*	15.418 (0.584)	3.18	*0.37 (0.02)*
	Lambs	401	377 (308)	113 (8)	NS	NS	*NS*	5.679 (3.103)	2.10	*0.13 (0.7)*	28.507 (4.346)	4.70	*0.66 (0.10)*	8.833 (2.531)	2.62	*0.20 (0.06)*
	Adults	1841	659 (436)	126 (9)	12.437 (3.004)	2.80	*0.32 (0.07)*	NS	NS	*NS*	11.954 (2.405)	2.75	*0.30 (0.06)*	14.968 (0.626)	3.08	*0.38 (0.02)*

Table 3.1. (Continued)

Trait	Dataset	No. records	No. animals (genotyped)	Mean (SE[a])	V_A[b] (SE[a])	CV_A[c] (%)	h^{2d} (SE[a])	V_M[e] (SE[a])	CV_M[f] (%)	$m^2 g$ (SE[a])	V_C[h] (SE[a])	CV_C[i] (%)	$c^2 j$ (SE[a])	V_E[k] (SE[a])	CV_E[l] (%)	$e^2 m$ (SE[a])
Hindleg length [mm]	All	2379	740 (512)	177 (12)	17.884 (4.873)	2.40	*0.25 (0.06)*	10.777 (2.259)	1.86[a]	*0.15 (0.03)*[a]	32.891 (4.101)	3.24	*0.47 (0.06)*	8.834 (0.329)	1.68	*0.12 (0.01)*
	Lambs	425	400 (329)	162 (10)	NS	NS	*NS*	9.897 (4.999)	1.95	*0.14 (0.07)*	55.045 (6.029)	4.59	*0.77 (0.07)*	6.831 (1.952)	1.62	*0.10 (0.03)*
	Adults	1954	659 (436)	180 (10)	25.421 (5.074)	2.80	*0.46 (0.08)*	NS	NS	*NS*	21.132 (3.618)	2.55	*0.38 (0.07)*	8.459 (0.337)	1.61	*0.15 (0.01)*
Body weight [kg]	All	1672	579 (407)	21.62 (6.45)	1.650 (0.699)	5.94	*0.12 (0.05)*	3.736 (0.929)[a]	8.94[a]	*0.26 (0.05)*[a]	4.323 (0.703)	9.62	*0.30 (0.05)*	4.559 (0.211)	9.88	*0.32 (0.03)*
	Lambs	401	380 (314)	13.78 (2.98)	NS	NS	*NS*	1.104 (0.511)	7.62	*0.20 (0.09)*	3.211 (0.578)	13	*0.59 (0.11)*	1.086 (0.356)	7.56	*0.20 (0.07)*
	Adults	1297	396 (228)	23.95 (5.20)	2.475 (0.948)	6.57	*0.23 (0.08)*	NS	NS	*NS*	4.282 (0.885)	8.63	*0.40 (0.08)*	3.963 (0.191)	8.31	*0.37 (0.03)*
Metacarpal length [mm]	Adults	450	449 (332)	78 (6)	7.204 (2.076)	3.44	*0.45 (0.11)*	NS	NS	*NS*	NS	NS	*NS*	8.762 (1.611)	3.79	*0.55 (0.11)*
Jaw length [mm]	Adults	566	565 (396)	112 (13)	7.824 (2.155)	2.48	*0.39 (0.10)*	NS	NS	*NS*	NS	NS	*NS*	12.113 (1.753)	3.09	*0.61 (0.10)*

[a] Standard error; [b] Additive genetic variance; [c] Coefficient of additive genetic variation; [d] Heritability; [e] Maternal genetic variance; [f] Coefficient of maternal genetic variation; [g] Maternal genetic effect [h] Permanent environmental variance; [i] Coefficient of permanent environmental variation; [j] Permanent environmental effect (ratio between permanent environmental variance and total phenotypic variation); [k] Residual variance; [l] Coefficient of residual variation; [m] Residual effect.

Two neonatal traits were considered. Daily observations allow us to identify *birth date* for each lamb to the nearest day, measured as the number of days from the 1st of January. New born lambs are captured, tagged and weighed to give *birth weight*; in this analysis we included measurements collected within four days of birth. Both birth date and birth weight were first analysed as traits of the lamb and second analysed as maternal traits. In the latter case the trait represented the ability of a dam to give birth to lambs on different dates or of different weights.

Sheep older than four months of age are regularly captured, especially during an annual August catch up but also at lower frequency during Spring and Autumn. In addition many animals are found within 24 hours of death. At each capture and at *post mortem*, *body weight* is obtained. Due to large seasonal variation, the body weight data analysed here was restricted to data collected in August. *Foreleg length* measured in mm from as the length of the metacarpal bone measured when both the knee joint and the hoof joint are bent away from it and *hindleg length* measured in mm from the tubercalcis of the fibular tarsal bone to the distal end of the metatarsus is also collected at each capture and *post mortem*. The genetic basis of variation in body weight and hindleg length varies over ontogeny (Wilson et al. in prep) and so these traits, together with foreleg length, were analysed in three different data subsets defined by the age of the animals included. The first (unrestricted) dataset included animals of all ages (minimum 47 days, maximum 15 years, mean 3.8 years); the second dataset included only those animals younger than nine months (referred to as lambs), and the third dataset only animals older than nine months (referred to as adults). This classification was applied since for the two leg length measures it appeared to separate the juvenile period when heritability is low but maternal

effects are strong from the adult period when heritability is high and maternal effects are no longer detectable.

Two additional measures of body size were available from cleaned skeletal material. *Metacarpal length* was measured as the distance between the proximal and distal canal foramina on the dorsal side of the metacarpus. *Jaw length* was measured as the distance between the gonion caudale and the most aboral indentation of the mental foramen. For both measures, all data were for adult animals (i.e. older than nine months).

Definition of Fixed Effects – Fixed effects known to influence the study traits were fitted in the variance components models (see below). Table 3.2 lists the effects fitted for each trait and reports the number of degrees of freedom used by each effect. A general linear model analysis implemented in Minitab 14.1 (Minitab Inc.) was applied to determine the fixed effects significantly contributing to variation in the study traits. Sex was fitted with two levels (male or female); litter size with either two (twin or singleton) or three levels (twin, singleton, unknown). The age of the mother was classified into 11 levels (one year old to ten or more, plus one level for unknown age). Although the age of the mother is a continuous variable, the fitting of mother's age as multilevel factor rather than as a covariate allowed better correction of the study trait and allowed the use of records where the age of the mother was unknown. Birth date was fitted as a covariate and was measured as days from the 1st of January of the year of birth. Birth year and capture year had one level for each year to control for differences in environmental conditions (e.g. population density) at the time of birth or measurement. Capture age accounted for the growth of the animal and was fitted as a covariate in lamb foreleg and hindleg length, in lamb body weight, and in birth weight as trait of the mother. In birth weight as trait of the offspring, capture age was fitted as factor with 4 levels, one for each

day from birth. In the other traits capture age was divided into 11 levels for age from 0 to 1 year (one level for each month) and 14 levels after the first year of age (one level for each year). Mother's cohort was fitted as a factor where each level was a different year of birth of the mother.

Table 3.2. Fixed effects for the study traits fitted in the polygenic and QTL models. Numbers are the degrees of freedom used by each effect (NF: not fitted).

Trait	Dataset	Sex	Litter size	Mother's age	Birth date	Birth year	Capture year	Capture age	Cohort
Birth date (lamb)	Neonatal	1	NF	10	NF	25	NF	NF	NF
Birth date (mother)	Neonatal	1	2	NF	NF	20	NF	NF	24
Birth weight (lamb)	Neonatal	1	1	10	NF	16	NF	3	NF
Birth weight (mother)	Neonatal	1	1	10	1[a]	19	NF	1[a]	NF
Foreleg length	All ages	1	2	NF	NF	24	17	24	NF
	Lambs	1	2	NF	NF	14	NF	1[a]	NF
	Adults	1	2	NF	NF	24	17	16	NF
Hindleg length	All ages	1	2	NF	NF	24	17	24	NF
	Lambs	1	2	NF	NF	14	NF	1[a]	NF
	Adults	1	2	NF	NF	24	17	16	NF
Body weight	All ages	1	2	NF	NF	23	17	16	NF
	Lambs	1	2	NF	NF	14	NF	1[a]	NF
	Adults	1	2	NF	NF	22	17	10	NF
Metacarpal length	Adults	1	2	NF	NF	22	NF	11	NF
Jaw length	Adults	1	2	NF	NF	24	15	15	NF

Variance Components Estimation – Under the null hypothesis of no segregating QTL (i.e. major genes whose effect stands out from the average genetic background), it is assumed that the additive genetic variation is represented by a number of genes with small effect randomly scattered across the genome, and for this reason this design is called the polygenic model. The polygenic model provides the log-likelihood against which to test the alternative hypothesis of linkage. In addition, it yields information about the relative weight of the different variance components on the total variation.

Under a polygenic model, fixed effects can be included to account for known influences on the phenotypic mean, while the remaining variance is partitioned among specified random effects (Lynch and Walsh, 1998; Williams and Blangero, 1999). In the simplest case the random effects will include just the additive genetic value such that:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

Where **y** is a vector of records on individuals; β is a vector of fixed effects, **a** is a vector of additive genetic effects (or breeding values) estimated on the basis of the coefficient of co-ancestry between any pair of individuals in the pedigree; **e** is a vector of residual effects. **X** and **Z** are design matrices relating records to the appropriate fixed or random effects. The appropriate fixed effects were determined separately for each trait. Additional components responsible for the total variation (random effects), such as permanent environment and maternal effect, were fitted if they significantly improved the likelihood of the models (likelihood ratio test, the test statistic under the null hypothesis that the variance component is zero is a 50:50 mixture of zero and a χ^2 with 1 df). Variance components for each trait are reported in Table 3.1. The additive genetic relationship matrix created from the pedigree file incorporated information from all known and inferred relatives, of both sexes, correctly weighted for relatedness. Where

different measurements of the same trait, on the same individual, were available at different life stages, the permanent environmental effect grouped the repeated measurements to determine the environmental variance between individuals that arose from long-term or non-localized conditions. Finally, the maternal effect removed variation due to the contribution of the mother's phenotype and genotype on the offspring's trait. It should be noted that if the maternal effect is not explicitly modelled, its variation is included in the additive genetic component thus leading to possibly biased results.

Heritability (h^2), maternal effect (m^2), permanent environment effect (c^2), and residual effect (e^2) were calculated as the ratio of the relative variance component (V_A, additive genetic variance; V_M, maternal genetic variance; V_C, permanent environmental variance; V_E, residual variance) to total phenotypic variance (V_P), i.e. $h^2 = V_A/V_P$; $m^2 = V_M/V_P$; $c^2 = V_C/V_P$; $e^2 = V_E/V_P$.

The coefficient of variation (CV) standardizes the variance by the trait mean instead of the total variance, and it is calculated as the ratio of the standard deviation (square root of the variance) to the mean times 100 therefore:

$$CV_i = 100V_i^{1/2} / \overline{x},$$

Where the subscript i stands for the additive genetic (A), maternal effect (M), permanent environment (C), and residual components (E) and $\overline{x}$ is the trait mean.

Variance components were estimated by the restricted maximum likelihood procedure (Lynch and Walsh, 1998) implemented in the software package ASReml (Gilmour et al., 2002). The estimated parameters of the variance components under the null hypothesis of no QTL effects are listed in Table 3.1.

QTL Mapping – To map putative segregating QTL, an IBD (identity by descent) matrix estimated at any given map position was fitted in the polygenic model described above as an additional random effect (George et al., 2000):

$$\mathbf{Y = X\beta + Za + Zq + e}$$

Where **q** is a vector of additive QTL effect. As the IBD matrix represents the allele sharing probability between any pair of individuals in a specific genomic region, the significance of its effect on the study trait is evidence of segregating QTL in the region being tested. IBD sharing statistics were estimated using pedigree relationships, marker data, and map distances described above in more detail in Beraldi et al. (2006). For an initial scan, IBD matrices and variance components were estimated every 10 cM. Putative QTL regions were then scanned every 1 cM. The IBD sharing analysis was performed by a Markov chain Monte Carlo (MCMC) procedure which is based on a stochastic process (gene-drop simulations) and as such does not provide an exact result, but allows the handling of very large and complex pedigrees. After a burn-in period of 1000 cycles, 100,000 MCMC iterations where performed and sample statistics were stored every ten iterations. This process was implemented in the program Loki (Heath, 1997). The IBD matrices were then inverted and fitted one by one in ASReml using a program written by one of the authors (AFM) to automate the process of inputting and storing the results. LOD scores were calculated as the difference in log-likelihood between QTL and polygenic model according to the equation:

$$LOD = (L_1 - L_0) / \ln(10)$$

Where L_1 is the natural log-likelihood of the QTL model and L_0 the natural log-likelihood of the polygenic model.

The significance thresholds adopted in this study to declare evidence of a QTL correspond to those suggested by Lander and Kruglyak (1995) for human pedigrees; this decision was taken on the basis that the size of the Soay sheep and human linkage maps are very similar (~3300 cM). The LOD value of 3.3 denotes the *genome-wide* significance, i.e. the probability of finding a false positive every 20 genome scans; the value 1.9 corresponds to the *suggestive* linkage which is the evidence expected to occur once at random in a genome scan (Nyholt, 2000). Confidence intervals for the presence of a putative QTL were defined by the map range within a one-LOD score drop from the peak value, this is equivalent to approximately 95% confidence (Lander and Botstein, 1989).

3.4. Results

Variance Components Analysis – Under the null hypothesis of no QTL effects, variance components analysis provides estimates of the population parameters. Consistent with Wilson et al. (2005a) we found that birth date had a low heritability (h^2= 0.07) as a trait of the lamb, but a substantial maternal effect (m^2 = 0.69) and when analysed as a trait of the mother a much higher heritability (h^2= 0.28, Table 3.1). Similarly birth weight had a lower additive genetic component than maternal component when analysed as a trait of the lamb (h^2= 0.16, m^2= 0.25) but a higher heritability as a trait of the mother (h^2= 0.27; Table 3.1). The coefficient of variation for birth date was not computed as the mean value of this trait is not meaningful since it is arbitrary when to start counting the days before birth.

With respect to body size traits measured on individuals older than neonates, the analysis of foreleg and hindleg length and body weight across all ages showed a moderate to low heritability (h^2 foreleg length = 0.16, h^2 hindleg length = 0.25, body weight = 0.12;

Table 3.1) while the maternal effect, fitted only for animals of up to nine months of age, explained 3% of the variation in fore leg length, 15% of the variation in hindleg length, and 26% of the variation in body weight (Table 3.1). When the data were restricted to measurements taken on lambs (<9 months of age), no additive genetic component was detected for any of the three measures of body size, but a maternal effect was found for all three traits (Table 3.1). When the data were restricted to measurements taken on adults (≥ 9 months of age), heritability was moderately high for leg length ($h^2 = 0.32$ for foreleg length and 0.46 for hindleg length) and moderate for body weight ($h^2 = 0.23$), and no maternal effect was detected (Table 3.1).

Both skeletal measurements had substantial heritabilities ($h^2 = 0.45$ for metacarpal length $h^2 = 0.39$ for jaw length; Table 3.1) and, consistent with the other traits when measured on individuals greater than nine months of age, no maternal effects were detected in these traits (Table 3.1).

Genome Scans – Genome scans were performed for the traits having an additive genetic component greater than zero; therefore, leg length and body weight in lambs were not investigated further. The LOD score profiles of the whole genome are shown in Figure 3.1 and 3.2.

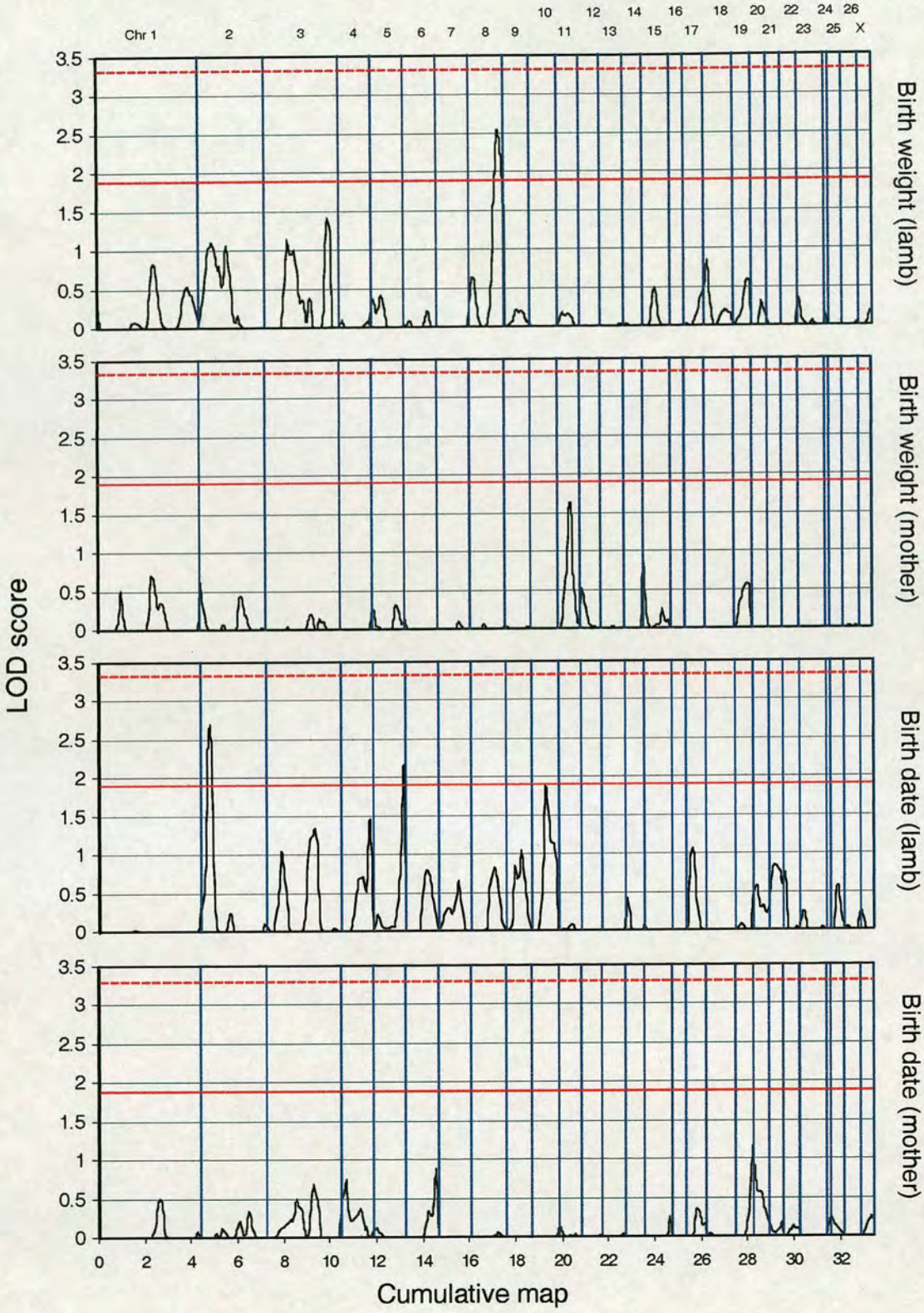

Figure. 3.1 – Whole genome scans of neonatal traits. Top to bottom: birth weight as a trait of the lamb and of the mother and birth date as a trait of the lamb and the mother. LOD score values (ordinates) were plotted against genetic position (abscissas, Morgan scale). Dotted lines show the genome-wide significance threshold (3.3); dashed lines are the suggestive significance threshold (1.9). Vertical lines mark the chromosome boundaries and chromosome names are displayed at the top.

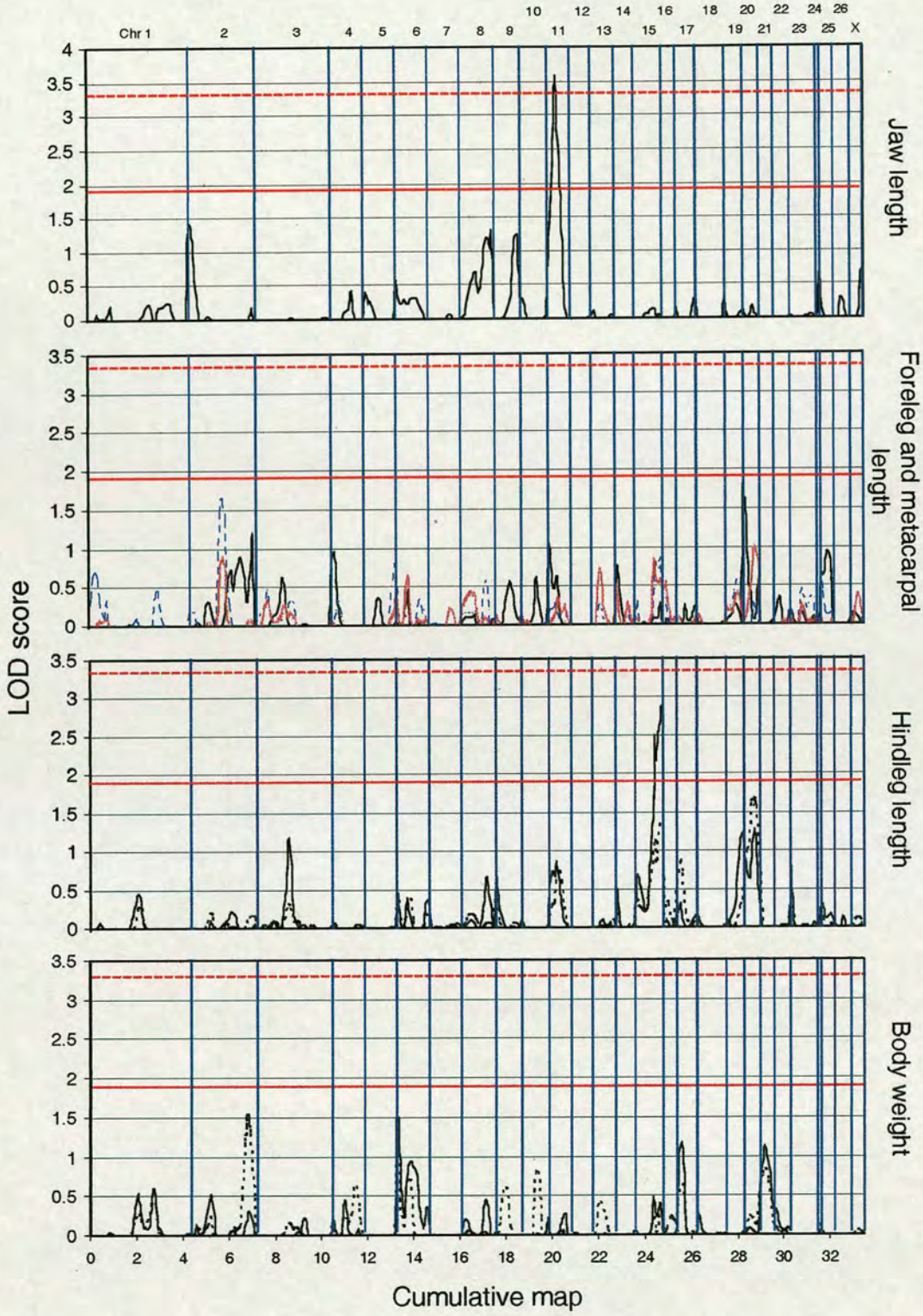

Figure 3.2 – Whole genome scans of body size traits. Top to bottom: jaw length, foreleg length (continuous line, metacarpal length; dotted line, adults; hatched line, animals of all ages), hindleg length (continuous line, adults; dotted line, animals of all ages) and body weight (continuous line, adults; dotted line, animals of all ages). Graph legend as in Figure 3.1.

The highest LOD score values identified for each trait are listed in Table 3.3 along with the variance components estimates and the map characteristics of the region containing the peaks. Suggestive evidence of a QTL was detected on chromosome 2 for birth date as a trait of the lamb (LOD= 2.70) and on chromosome 8 for birth weight as a trait of the offspring (LOD= 2.54). In both cases the suggestive QTL accounted for all of the additive genetic variation. Chromosome 15 showed suggestive evidence of a QTL for hindleg length in adults (LOD= 2.89) which explained 36% of the variation (residual genetic variation: h^2= 0.09). Finally, a genome wide significant QTL was detected for adult jaw length on chromosome 11 (LOD= 3.59). The IBD matrix corresponding to the position harbouring the QTL explained 29% of the phenotypic variation, thus reducing the residual genetic variation to h^2= 0.15. The LOD profiles of these chromosomes are shown in more detail in Figure 3.3. QTL confidence intervals derived from the 1-LOD drop support correspond to approximately 30 cM for birth date (lamb), 40 cM for birth weight (lamb), 40 cM for hindleg length and 20 cM for jaw length (Figure 3.2, Table 3.3).

Table 3.3. Highest QTL LOD scores for the study traits and their estimated parameters

Trait	Dataset	LOD	Chr.	Position (cM)	Flanking markers (cM)[a]		1-LOD drop support (cM)	q^2 (SE)[b]	h^2 (SE)	m^2 (SE)	c^2 (SE)
Birth date (lamb)	Neonatal	2.70*	2	53	LPLP2 (12)	McM505 (1)	30	0.14 (0.04)	0.00 (0.00)	0.70 (0.03)	-
Birth date (mother)	Neonatal	1.16	20	1	BM1815 (1)	OLADRB2 (24)	-	0.34 (0.08)	0.00 (0.00)	-	0.10 (0.06)
Birth weight (lamb)	Neonatal	2.54*	8	130	URB024 (16)	BMS1967 (8)	40	0.21 (0.07)	0.00 (0.00)	0.26 (0.06)	-
Birth weight (mother)	Neonatal	1.63	11	61	MCM120 (3)	ETH3 (10)	-	0.26 (0.03)	0.00 (0.00)	-	-
Fore leg length	All	0.99	20	54	BM1818 (2)	BM1905 (18)	-	0.12 (0.07)	0.06 (0.08)	0.03 (0.02)	0.43 (0.05)
	Adults	1.65	2	145	BM81124 (12)	CP79 (10)	-	0.22 (0.10)	0.10 (0.12)	-	0.30 (0.06)
Hindleg length	All	1.69	20	45	OMHC1 (10)	BM1818 (12)	-	0.21 (0.08)	0.07 (0.10)	0.15 (0.03)	0.44 (0.06)
	Adults	2.89*	15	113	BMS2076 (13)	MCM105 (6)	40	0.36 (0.11)	0.09 (0.13)	-	0.39 (0.07)
Body weight	All	1.55	2	253	BM6444 (0)	BM2113 (26)	-	0.13 (0.04)	0.00 (0.00)	0.26 (0.05)	0.29 (0.04)
	Adults	1.50	6	1	BM9058 (1)	MCM0204 (4)	-	0.22 (0.07)	0.00 (0.00)	-	0.41 (0.07)
Metacarpal length	Adults	1.82	20	7	BM1815 (7)	OLADRB2 (18)	-	0.45 (0.10)	0.00 (0.00)	-	-
Jaw length	Adults	3.59**	11	37	CSSME70 (9)	SRCRSP6 (1)	20	0.29 (0.10)	0.15 (0.13)	-	-

[a] In parentheses the distance (cM) of the flanking markers from the QTL peak; [b] QTL heritability (QTL variance/total phenotypic variance); * LOD score chromosome wide significant (LOD> 1.9); ** LOD score genome wide significant (LOD> 3.3)

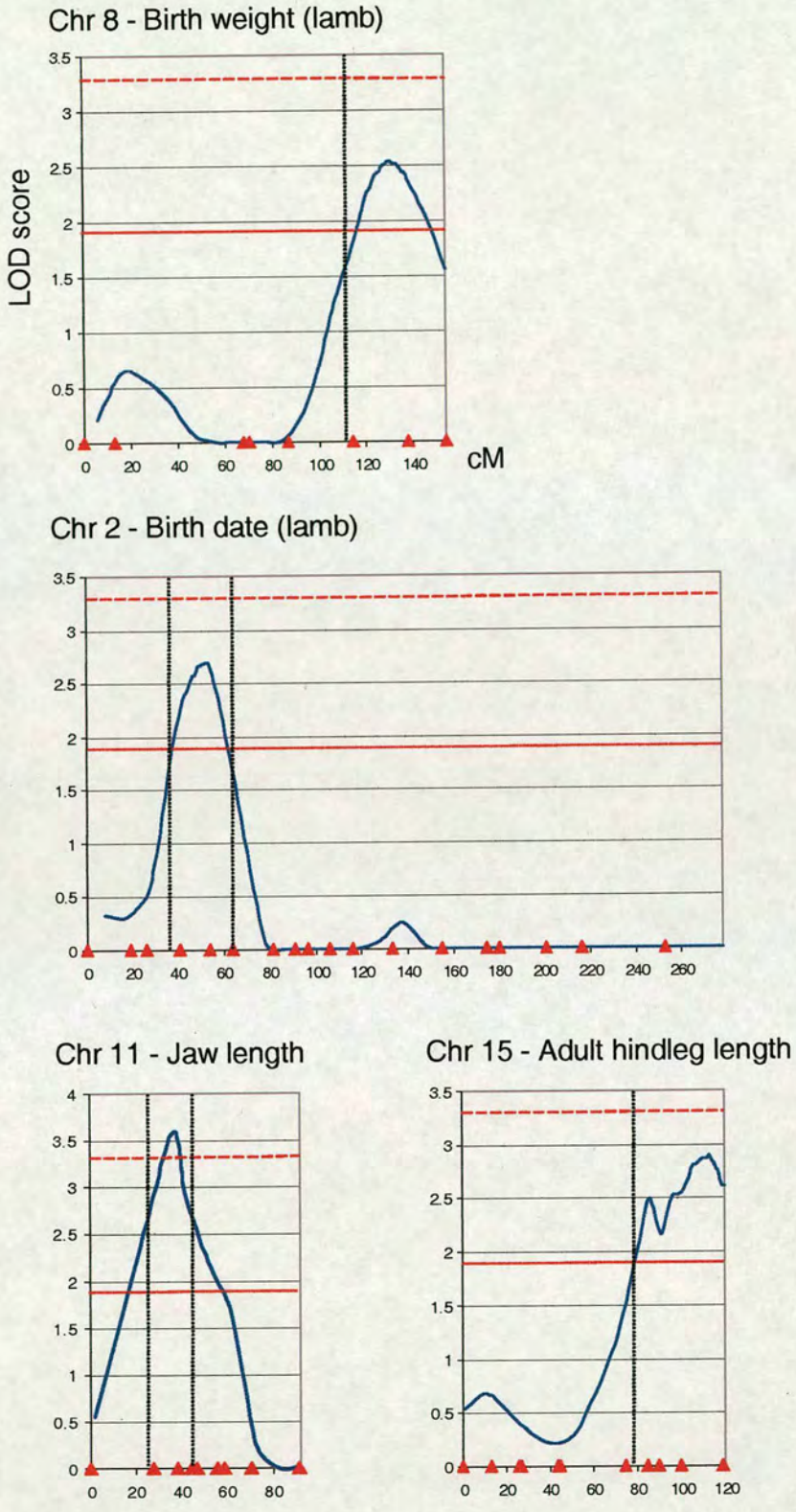

Figure 3.3 – Detailed map positions of the candidate QTL identified by the genome scans. Vertical lines define the 1-LOD drop confidence intervals. Triangles on the abscissas (chromosome maps) show the marker positions. Dotted and dashed horizontal lines represent the suggestive and significant linkage thresholds as in Figure 3.1.

3.5. Discussion

This article reports the results of genome scans aimed at identifying QTL responsible for fitness-related traits in a free living animal population, the Soay sheep on Hirta, St. Kilda. Using the same pedigree of animals and linkage map, a previous study has identified the loci underlying polymorphisms in coat colour and horn type, traits also correlated with fitness but encoded by single loci with Mendelian inheritance (Beraldi et al. 2006; Gratten et al. In press). Here we used QTL mapping to investigate traits with a continuous phenotypic distribution shaped by both genetic and environmental factors. Such quantitative traits are probably the more likely raw material for microevolutionary change. The traits targeted in this project represent different aspects of early development and body size, and several of them have documented relationships with total fitness (see Introduction) (Clutton-Brock et al., 1992; Milner et al., 2000; Jones et al., 2005). QTL mapping projects in natural populations are still at a pioneering stage, but this study provides support for the idea that QTL of large effect may be detectable by this approach.

Variance Components QTL Analysis – Different statistical and experimental strategies have been devised for mapping QTL using crosses of inbred or outbred lines of animal or plant populations (Haley et al., 1994; Zeng, 1994; Lynch and Walsh, 1998). However, these methods are less suitable for QTL mapping in large complex pedigrees of animals because they are designed for analysis of relatively simple pedigrees or single unrelated families, and consequently do not fully exploit the information of the range of pedigree relationships. The mapping panel used in this study presents a probably unprecedented level of complexity for a QTL genome scan. Furthermore, in common with many long term studies of natural populations, missing information is present in the

Soay sheep pedigree, genotype and phenotype files. All of the 882 pedigree members were related to each other, although they belonged to several different half-sib families and spanned up to seven generations including many inbreeding loops (175 loops detected by Loki). Such a level of complexity may not be unusual in natural populations although similar studies of QTL mapping are still rare (Slate et al., 2002; Slate, 2005). The variance components method chosen to perform the genome scans in Soay sheep was originally designed for the analysis of complex traits in human datasets (Almasy and Blangero, 1998; Williams and Blangero, 1999), where missing and unbalanced data are also common, but pedigree files are much simpler than in the Soay sheep. Following George et al. (2000), we performed a two-step variance components analysis where the first step involved the estimation of IBD matrices using an MCMC procedure. The MCMC sampling method allowed the handling of the entire pedigree in order to exploit the genetic relationship between all the possible pairs of individuals. The second step combined the phenotypic, pedigree and IBD information in a REML framework which is relatively robust to deviations from assumptions such as missing or unbalanced data and departure from normality.

Neonatal Traits – Using the full Soay sheep pedigree, Wilson et al. (2005a) found that in Soay sheep birth date and birth weight are determined more by the maternal performance (maternal effect) than by the offspring. Consistent with this observation, in the mapping panel we found a heritability of 0.07 and 0.28 as a trait of the lamb and trait of the mother, respectively (Table 3.1). Likewise, we found a heritability of 0.16 and 0.27 for birth weight as a trait of the lamb and trait of the mother, respectively (Table 3.1). Therefore, we scanned the genome to detect QTL affecting birth weight and birth date as a trait of both the lamb and the mother. We found a suggestive linkage (LOD= 2.70) for birth date as a trait of the lamb on chromosome 2 and a suggestive linkage (LOD=

2.54) for birth weight as a trait of the lamb on chromosome 8. A previous study in a commercial sheep breed did not examine birth weight as a trait but identified suggestive QTL on chromosome 2, 3 and 18 for weight at eight weeks of age (Walling et al., 2004), but these results had limited significance and could not be replicated (Johnson et al., 2005). In cattle, suggestive QTL for birth weight have been detected on chromosome 21 and 26 (Casas et al., 2003; Casas et al., 2004) which are homologous to chromosome 4 and 18, respectively, in sheep. At the moment, we are not aware of published genome scans performed to identify QTL for birth date in sheep or cattle.

The scans for birth date and birth weight as traits of the mother did not produce any particular evidence for QTL, the highest LOD score being 1.16 and 1.63 respectively. This is perhaps a reflection of the fact that the power of analysis was reduced by the small number of genotyped mothers (n= 136 for birth date, n=133 for birth weight) compared with genotyped offspring (n= 526 for birth date, n= 507 for birth weight; Table 3.1). To the best of our knowledge, no previous studies have been undertaken to detect QTL affecting birth date or birth weight as traits of the mother in sheep.

Body Size Traits – The phenotypic datasets for foreleg, hindleg length and body weight were initially analyzed including animals of all ages and then re-analyzed to include only lambs or only adults. This strategy was pursued to define traits with a higher proportion of genetic variance, although it also resulted in a reduction of the sample size. In the pedigree analyzed in this study, there was no additive genetic variation for either leg length measure or body weight in lambs, whereas adults had significant heritable variation for leg length and body weight. Overall, this trend suggests that the same trait has different sources of additive genetic or environmental variation at different ontogenic stages. Body size, for example, was more strongly affected by the permanent

environmental effect in lambs than in adults, especially with respect to hindleg length (hindleg length c^2= 0.77 in lambs, c^2= 0.38 in adults, c^2= 0.47 in all ages combined; Table 3.1). This pattern of change in variance components over life histories has been more thoroughly investigated by Wilson et al (in prep) using random regression on the full Soay sheep dataset, who also found little additive genetic variance (either direct or maternal genetic) in weight and hindleg length in lambs captured at four months of age.

In general, the linear traits (leg lengths, metacarpal and jaw length) had high values of h^2 and low values of CV_A whereas body weight had lower h^2 but higher CV_A (Table 3.1). To make a specific comparison, although the heritability of adult hindleg length was found to be more than twice that of adult body weight, the coefficient of variation was higher for body weight than for hindleg length (CV_A= 2.80 vs. 6.57). With respect to the permanent environmental effect, V_C had a similar value in both traits in terms of the proportion of total variation explained (c^2= 0.38 hindleg length vs. 0.40 body weight). However, the coefficient of variation of V_C for body weight was more than three times that of hindleg length (CV_C= 2.55 vs. 8.63). This trend confirms the intuition that body weight is composed of several underlying traits which confer a higher overall genetic contribution to the trait, but also a higher residual variation which in turn results in a lower ratio V_A/V_P (i.e. heritability).

Genome scans were conducted on all body size traits except in lamb data sets (animals of age up to nine months) due to the absence of heritable variation (see above). No QTL exceeding statistical thresholds for significance were detected for foreleg length or metacarpal length. In adults, the most significant result for hindleg length was identified on chromosome 15 (LOD= 2.89 Table 3.3). Unfortunately, the scan performed by Walling et al. (2004) in commercial sheep did not include chromosome 15

so that it is not possible to compare the results. Hindleg length in all ages combined produced the highest LOD score on chromosome 20 (LOD= 1.69, Table 3.3) where putative QTL for fat depth have been recorded (Walling et al., 2004). The same region harbours the MHC, a gene complex involved in the immune response and affecting parasite resistance in Soay sheep (Paterson et al., 1998). At this stage, the statistical significance of the result does not allow in depth speculation, but further research in this direction may uncover a relationship between the parasite resistance conferred by the MHC region and body size. The two highest LOD score peaks identified for body weight were located on chromosome 2 (LOD= 1.55, animals of all ages) and 6 (LOD= 1.50, adults, Table 3.3). The peak on chromosome 6 corresponds to the region identified by Walling et al. (2004) for muscle depth. The genome scan for jaw length produced a genome wide significant QTL with LOD score of 3.59. This trait represents a component of body size and, as such, is probably under selection (Milner et al., 1999). As far as we are aware, no other studies of this trait have been conducted in ruminants.

Improved modelling of traits for which multiple measures are available at different ages, such as body size, could be achieved by applying a random regression analysis in order to allow for the change of the genetic effect over time (Macgregor et al., 2005). We pursued this strategy by fitting in the variance component model the first degree polynomial of the QTL effect. Although the random regression model was no better than the model with constant QTL effect (results not shown), this does not preclude the possibility that better modelling of longitudinal traits may improve the performance of the analysis.

Effect Sizes – Although we report estimated effect sizes in Table 3.3 in order to provide a comparison with similar studies these estimates are inevitably inflated. In fact,

the estimation of the QTL heritability and effect from a genome scan aimed at mapping is upwardly biased and probably unrealistic, especially when the QTL has weak effect (Goring et al., 2001). This is because the maximum likelihood procedure provides the highest statistical evidence of linkage (LOD score) where the parameters (QTL effect) are maximized. Therefore, independent datasets should be analysed if both QTL position and effect are desired (Goring et al., 2001). Furthermore, in the case of studies focused on natural selection, a case can be made that effect sizes should be estimated without fitting fixed effects in the analysis, since selection only 'sees' raw variation.

Future Developments – We have performed one of the first genome wide scans in a free-living animal population in order to investigate the genetics of fitness-related traits and demonstrated that QTL peaks exceeding threshold criteria can be found. One genomic region is likely to carry genes of biological and evolutionary relevance and three other regions have suggestive QTL. The key challenge now is to follow up this work to confirm the validity of the putative QTL detected. Genotyping of additional markers around the putative QTL documented here in the existing pedigree will be conducted in order to improve the estimates of the IBD sharing probabilities and hence the resolution of the confidence intervals and to confirm or reject the hypothesis of linkage. Secondly, an independent set of families genotyped at the putative QTL regions will allow confirmation of the presence of QTL and more realistic estimation of QTL effect sizes.

Previous simulation studies have established that linkage disequilibrium in Soay sheep declines rapidly with genetic distance and that the overall linkage disequilibrium background is low (McRae et al., 2005). This pattern of linkage disequilibrium should allow the fine mapping of QTL through association analysis if the target region can be enriched with markers to a resolution equal to or less then 2cM/marker (McRae et al.,

2005). Such a marker density could be achieved combining microsatellites and SNP markers and by typing a larger number of individuals than were used here. We have previously managed to fine map Mendelian trait loci by linkage disequilibrium mapping (Gratten et al. In press) and similar efforts are underway to fine map quantitative traits.

Several other long term studies of individuals living in the wild could soon be in a position to conduct QTL searches like ours. In many cases large pedigrees have been recovered from a mixture of behavioural observations and DNA profiling, sufficient for quantitative genetic analysis (Kruuk, 2004). In addition the number of markers available to create genetic maps in non-model organisms is also steadily rising (e.g. Hansson et al., 2005) and thanks to the agricultural and human genetics research effort, the statistical challenge of searching for QTL within complex unmanaged pedigrees can be met. Here, we have presented our results as fully as possible, including all null results, in order to provide comparison for future studies.

Acknowledgements

We thank JG Pilkington and many previous volunteers and project members for collecting field data and genetic samples. We thank L. Buchanan, A. Stripe, C. Donald and J. Birch who measured skeletal material and A. Wilson for statistical advice and for comments on the manuscript. We thank the National Trust for Scotland for granting permission to work on St. Kilda, and QinetiQ for logistical support. The long term data collection on St. Kilda has been supported by Natural Environment Research Council (NERC) and Wellcome Trust grants to T.H. Clutton-Brock, B.T. Grenfell, L.E.B. Kruuk, M.J. Crawley and JMP. This study was funded by NERC through its Environmental Genomics thematic programme (grant number NER/T/S/2002/00189).

Chapter 4.

Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*)

Dario Beraldi[1,4], Allan F. McRae[1,2,5], Jacob Gratten[3,6], Jon Slate[3,7], Peter M. Visscher[1,2,8] and Josephine M. Pemberton[1,9]

[1] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, UK.*

[2] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, 4029, Australia.*

[3] *Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK.*

[4] dario.beraldi@ed.ac.uk; [5] allan.mcrae@qimr.edu.au; [6] j.gratten@sheffield.ac.uk;

[7] j.slate@sheffield.ac.uk; [8] peter.visscher@qimr.edu.au; [9] j.pemberton@ed.ac.uk

Author contribution: DB performed the statistical analysis, and wrote the manuscript. AFM and PMV helped with the statistical analysis. All the contributing authors critically read and improved drafts of this manuscript. The whole project was planned and supervised by JMP, JS and PMV.

4.1. Abstract

A genome wide scan was performed to detect quantitative trait loci (QTL) for resistance to gastrointestinal parasites and ectoparasitic keds segregating in the free-living Soay sheep population on St. Kilda (UK). The mapping panel consisted of a single pedigree of 882 individuals of which 588 were genotyped. The Soay linkage map used for the scans comprised 251 markers covering the whole genome at average spacing of 15 cM. The traits here investigated were the strongyle faecal egg count (FEC), the coccidia faecal oocyst count (FOC), and a count of keds (*Melophagus ovinus*). QTL mapping was performed by means of variance component analysis so that the genetic parameters of the study traits were also estimated and compared with previous studies in Soay and domestic sheep. Strongyle FEC and coccidia FOC showed moderate heritability (h^2= 0.26 and 0.22 respectively) in lambs, but low heritability in adults (h^2<0.10). Ked count appeared to have very low h^2 in both lambs and adults. Genome scans were performed for the traits with moderate heritability, and two genomic regions reached the level of suggestive linkage for coccidia FOC in lambs (LOD= 2.68 and 2.21 on chromosome 3 and X, respectively). This is the first study to report a QTL search for parasite resistance in a free-living animal population and therefore may represent a useful reference for similar studies aimed at understanding the genetics of host parasite co-evolution in the wild.

Keywords: QTL mapping; Soay sheep; Nematode; Variance components; Natural population

4.2. Introduction

The antagonism between host and parasite is thought to be a major force in ecology and evolution due to its potential to generate and maintain genetic variation. Parasites are often characterized by high potential for diversification due to their high speed of speciation (Dykhuizen, 1998), whereas the hosts they colonize constitutes a rapidly changing environment (Huyse et al., 2005). As a consequence, the host-parasite relationship generates continuously evolving host and parasite lineages (Nadler, 1995). In principle, this continuous battle can maintain genetic diversity in the antagonistic populations provided that a specificity between host and parasite genotypes is present (Haldane, 1949). Parasite resistance is likely to be controlled by several loci and therefore it may receive a strong mutational input which generates genetic variation (Houle et al., 1996). Host-parasite co-evolution may maintain genetic variation if the additive genetic value of a host genotype changes when parasites evolve as a response to the selection induced by the host (Haldane, 1949). Antagonistic pleiotropy may result in maintenance of genetic variation if the same genotype is positively selected for one fitness-related trait but negatively selected for another fitness-related trait (Roff and Mousseau, 1987). In the case of parasite resistance, this last hypothesis can occur since it has been found that sheep that are genetically resistant to intercellular infections may be more susceptible to infection from intracellular pathogens (Gill et al., 2000).

In sheep and other domestic ruminants, gastrointestinal nematodes are one of the most important classes of parasite. Since parasitic nematodes cause a loss of production in domestic sheep, intensive treatments based on anthelminthic drugs have been applied to control their infections. However, this practice is resulting in the widespread evolution of nematode strains tolerant to the drugs (Prichard, 1994; Kaplan, 2004). Intensive effort, therefore, has been invested in understanding, and exploiting through breeding

programs, the genetic basis of parasite resistance and host-parasite co-evolution (Kaplan, 2004). Parasite resistance is complex in nature having polygenic and environmental components (Stear et al., 1997; Bishop and Stear, 2003). Resistance to infection by gastrointestinal nematodes has moderate heritability in domestic sheep ranging from 0.13 (McEwan et al., 1992) to 0.53 (Baker et al., 1991) and resistant or susceptible lines have been selected in various countries (Dominik, 2005).

It would be of great interest, and have applied benefits, to know whether the documented genetic variation for parasite resistance is conferred by a few genes of major effect or due to a large number of genes of small effect. One strategy to dissect and understand the genetics of the host response to parasite infection is to locate the regions harbouring the genes responsible for the variation on a genetic map. Quantitative trait loci (QTL) mapping can help to identify candidate genomic regions and estimate their relative contribution to trait variation. To this end, different linkage mapping projects have been undertaken to find QTL for parasite resistance (Dominik, 2005). A genome scan was performed by Beh et al. (2000) using lines of sheep diverging for parasite resistance. Different regions were detected as likely to carry genes for resistance although no region was statistically significant after correcting for multiple tests. Davies et al. (2006) genotyped naturally infected lambs to scanned regions previously identified as candidates for either genes for resistance or genes for other economical traits could be replicated in their independent dataset. Evidence of linkage was found on chromosome 2, 13, 14 and 20. Recently, a genome scan performed by Crawford et al. (2006) (Crawford et al., 2006) using divergent lines and naturally infected animals detected a significant QTL on chromosome 8. In general, such selected populations have the advantage of high quality pedigrees and phenotypic data in terms of sample size and accuracy. In order to increase the power of analysis, the populations used for mapping purposes are usually

grown under controlled and uniform conditions designed to maximize the genetic contribution to the phenotype (Lymbery, 1996; Lynch and Walsh, 1998). Such experimental designs, however, may fail to capture those interactions between genes and environment that occur in natural populations (Erickson et al., 2004; Slate, 2005) and that could contribute to host parasite co-evolution and host population dynamics (Gulland and Fox, 1992; Gulland et al., 1993; Hudson et al., 1998).

From an evolutionary perspective, it is of interest to know whether major genes for parasite resistance explain observed variation in natural populations *in situ*. Perhaps such genes can only be detected under highly controlled environmental conditions when genetic variation and statistical power are maximised. In free-living populations, environmental noise and interactions between genetic variation and environmental variation may mask the effects of individual genes (Lynch and Walsh, 1998). In this paper we present a QTL analysis of parasite resistance in a free-living sheep population.

The free-living Soay sheep population on Hirta, St. Kilda UK, is the subject of a long term project aimed at addressing a wide range of ecological and evolutionary issues (extensively documented in Clutton-Brock and Pemberton, 2004) including the genetics and evolution of parasite resistance in the wild. The population dynamics of Soay sheep are characterized by periodic fluctuations in the number of individuals: the population size increases until the density of animals exceeds the winter carrying capacity and, as a result, a large proportion of individuals die in the following winter due mainly to starvation. High mortality is exacerbated by heavier burdens of nematodes which reduce the over-winter survival of infected host (Gulland et al., 1993; Wilson et al., 2004). The Soay sheep is naturally parasitized by several gastrointestinal nematode species (Wilson et al., 2004; Wimmer et al., 2004; Craig et al., 2006b), the most prevalent and abundant being strongyles (of which the predominant species are *Teladorsagia circumcincta*,

Trichostrongylus axei, *Trichostrongylus vitrinus*; see Craig et al. 2006). Different species of protozoans also infect the intestinal tract of Soay sheep; these belong mainly to the genus *Eimeria* but *Cryptosporidium parvum* and *Giardia duodenalis* also occur (Wilson et al., 2004; Craig et al., 2006a). It has been found that the overall burden of nematodes decreases and the relative abundance of different species changes with the age of the host (Craig et al., 2006b). The prevalence, intensity and diversity of protozoan species also tends to decrease with host age (Craig et al., 2006a). Keds (*Melophagus ovinus*) also parasitize the Soay sheep living in the wool and feeding on blood, causing anaemia and irritation.

The evolution of parasite resistance in Soay sheep has previously been addressed using two different population genetics strategies, the first being a quantitative genetics approach. Parasite resistance, measured as strongyle faecal egg count (FEC), is under directional selection (Coltman et al. 1999). In addition, there is a positive genetic correlation between body traits and resistance to strongyles (Coltman et al., 2001a) so that resistant sheep also experience better growth. Because body size and parasite resistance are under directional selection, it is expected that in this population the allelic variants associated with small body size and/or low parasite resistance will be eliminated by selection, and additive genetic (heritable) variation will be reduced to near zero (Fisher, 1958; Endler, 1986). However, parasite resistance in Soay sheep has low but not null heritable variation, previous population-wide estimates based on an animal model found a heritability for FEC in summer of 0.11 ± 0.02 in males and 0.13 ± 0.01 in females (Coltman et al., 2001a).

In a second approach, previous studies in Soay sheep have examined a number of candidate loci for parasite resistance in simple association studies. An allele of the interferon gamma gene (IFNG) on chromosome 3 was found to be associated with reduced strongyle FEC in four and 16 month old sheep (Coltman et al., 2001b) and with

increased antibody specific to *Teladorsagia circumcincta* in lambs. Paterson et al. found that some alleles at the major histocompatibility complex (MHC) region on chromosome 20 are associated with low survivorship and high levels of strongyle FEC or *vice versa* (Paterson, 1998; Paterson et al., 1998), although the FEC association was only apparent as an interaction with body weight. These observations were consistent with the theory that MHC variation is maintained by frequency-dependent selection (Paterson, 1998; Paterson et al., 1998). Finally, the three genotypes of the allozyme adenosine deaminase (ADA), an enzyme involved in purine metabolism and immune function, appear to be associated with over winter mortality and nematode burdens (Gulland et al., 1993). Genetic variation at the ADA locus seems to be maintained by the phenomenon of overdominance (heterozygote advantage) since the heterozygote individuals appear to experience less parasitic infection than the two homozygote genotypes (Gulland et al., 1993). It is of great interested to know whether these association studies can be supported in a more rigorous QTL search.

Here, we make use of a previously established mapping pedigree and linkage map (Chapter 2 and 3) and phenotypic data for three classes of parasites (gastrointestinal nematodes coccidia and keds), to ask (1) whether we can detect heritable variation for resistance (2) whether we can detect QTL for resistance and (3) whether any QTL found coincide with previous domestic sheep or Soay sheep studies.

4.3. Materials and Methods

Study population – The Soay sheep on the islands of Soay and Hirta (St. Kilda archipelago, North West Scotland, UK, 57°49' N, 08°34W) are feral populations of a breed regarded as the most primitive in Europe (Campbell, 1974; Doney et al., 1974);

nowadays, the sheep population of Hirta varies between 600 and 2000 individuals. Since 1985 regular expeditions have been sent to St. Kilda to monitor the population dynamics and to record the life histories of individuals living in Village Bay, Hirta (Clutton-Brock et al., 2004a). No predators are present on St. Kilda.

Mapping pedigree and linkage map – The whole Soay sheep pedigree file numbers more than 3900 animals. Within this pedigree maternal links were assigned through observation of the animals in the field, whereas paternal links were inferred through molecular analysis (Overall et al., 2005). From the total pedigree, a panel of 588 animals was genotyped at 247 microsatellite and four isoenzyme markers. This subset comprised all the half-sibships with ten or more individuals and their common parents. The ancestors of the genotyped individuals and the animals linking different sibships (n= 294) were not genotyped, but they were included in the mapping pedigree to improve the estimates of kinship and the identity by descent (IBD) coefficients in the variance component analysis. A more thorough description of the mapping pedigree and selection criteria is included in Beraldi et al. (2006). The Soay sheep map covers approximately 90% of the genome with an average inter-marker spacing of 15 cM. Further details of the map characteristics and of the technical procedures can be found in Beraldi et al. (2006).

Phenotypic dataset and measures of parasitism – Phenotypic records of the animals in the mapping pedigree were retrieved from the Soay sheep database. The data analysed in this study were collected between 1988 and 2005 from animals born between 1979 and 2002.

In the present study, the quantification of sheep resistance to gastrointestinal parasites was based on the indirect measures of strongyle faecal egg counts (FEC) and coccidia faecal oocyst count (FOC). The direct count of parasites would involve the

sacrifice of animals and post-mortem examination: this alternative is not feasible because the Soay sheep are protected and the sacrifice of animals would be in conflict with the study of the Soay sheep as a free-living population. However, previous work has shown a correlation between FEC and burden in island Soay populations (Wilson et al., 2004). Strongyle FEC and coccidia FOC was determined as the density of parasite eggs (FEC) or oocysts (FOC) per gram (wet weight) of faeces using a modification of the McMaster technique (MAFF, 1986). Other distinctive helminth species (*Nematodirus* spp., *Moniezia expansa*, *Capillaria longipes* and *Trichuris ovis*) are routinely classified and quantified in Soays but were not abundant enough for analysis. A few individuals that had previously been treated with either anthelminthics or hormones for experimental purposes were excluded from analysis. The count of keds was the total number of keds observed during a one minute search of the wool on the sheep's belly. The raw data (strongyle FEC, coccidia FOC, and ked count) were transformed into the natural logarithm to achieve a distribution closer to normality. The genetic and environmental sources of variation of parasite resistance are expected to change with the age of the animals and time of the year (Bishop et al., 1996; Coltman et al., 2001b). Therefore, only the samples collected in the August catch up, when most of the data are collected, were included in the analyses. In addition, each parasitic group was analysed separately in lambs (four month old animals) and adults (16 month old animals or older). Sample sizes and summary statistics for each trait are reported in Table 4.1.

Table 4.1. Characteristics and estimated variance components of the study traits.

Trait	Dataset	No. records	No. animals (genotyped)	Mean (SE)[a]	V_A[b] (SE)[a]	CV_A[c] (%)	h^2[d] (SE)[a]	V_C[e] (SE)[a]	CV_C[f](%)	c^2[g] (SE)	V_E[h] (SE)[a]	CV_E[i] (%)	e^2[j] (SE)[a]
Strongyles	Lambs	383	381 (307)	5.376 (2.150)	0.944 (0.472)	18.08	0.26 (0.12)	NS	-	-	2.626 (0.441)	30.14	0.74 (0.12)
Coccidia	Lambs	230	228 (204)	7.979 (1.228)	0.254 (0.254)	6.32	0.22 (0.21)	NS	-	-	0.915 (0.243)	11.99	0.78 (0.21)
Keds	Lambs	376	374 (310)	1.221 (0.804)	0.022 (0.054)	12.06	0.04 (0.09)	NS	-	-	0.563 (0.067)	61.45	0.96 (0.09)
Strongyles	Adults	962	345 (192)	3.073 (2.749)	0	-	-	0.789 (0.236)	28.90	0.13 (0.04)	5.479 (0.311)	76.17	0.87 (0.04)
Coccidia	Adults	694	240 (155)	5.224 (2.517)	0.310 (0.183)	10.66	0.06 (0.03)	NS	-	-	4.890 (0.305)	42.33	0.94 (0.03)
Keds	Adults	1303	396 (229)	0.123 (0.312)	0.002 (0.002)	40.97	0.03 (0.02)	NS	-	-	0.089 (0.004)	242.50	0.97 (0.02)

[a] Standard error; [b] Additive genetic variance; [c] Coefficient of additive genetic variation; [d] Heritability; [e] Permanent environmental variance; [f] Coefficient of permanent environmental variation; [g] Permanent environmental effect (ratio between permanent environmental variance and total phenotypic variation); [h] Residual variance; [i] Coefficient of residual variation; [j] Residual effect.

Definition of fixed effects – Fixed effects influencing the study traits were fitted in the variance component models. In order to facilitate comparisons with previous studies in Soay sheep, the fixed effects fitted for strongyle FEC were the same as those fitted by Coltman et al. (2001b). For consistency, coccidia FOC was also analyzed with the same model. A general linear model analysis implemented in Minitab 14.1 (Minitab Inc.) was applied to determine the amount of variation explained by each fixed effect (Table 4.2). Sex was fitted with two levels (male or female). Litter size was fitted as a three level factor (singleton, twin, or unknown). Birth year and collection year had one level for each year to control for differences in environmental conditions (e.g. population density) at the time of birth or measurement. Weight and age at collection were fitted as covariates with one degree of freedom.

Table 4.2. Fixed effects for the study traits fitted in the polygenic and QTL models. The deviance explained (in percentage) and the degrees of freedom used by each term (in brackets) are reported.

Trait	Dataset	Sex	Litter size	Birth year	Collection year	Weight	Collection age	Total deviance explained (df)
Strongyles	Lambs	4.2 (1)	0.3 (2)^{NS}	-	17.4 (14)	3.5 (1)	NF	25.5 (18)
Coccidia	Lambs	0.7 (1)^{NS}	~0 (2)^{NS}	-	20.8 (9)	5.6 (1)	NF	27.4 (13)
Ked count	Lambs	5.8 (1)	3.1 (2)	-	4.1 (14)^{NS}	NF	NF	13.0 (17)
Strongyles	Adults	13.2 (1)	NF	NF	4.0 (17)	0.4 (1)	NF	21.2 (22)
Coccidia	Adults	1.4 (1)	NF	NF	13.4 (12)	1.2 (1)	2.1 (1)	18.2 (15)
Ked count	Adults	NF	NF	4.1 (22)	5.8 (16)	NF	NF	9.9 (38)

^{NS}: Effect non-significant ($p>0.05$) but fitted for consistency with previous analyses.

Estimation of variance components – The model under the null hypothesis is a polygenic model which assumes that additive genetic variation is from a large number of genes with small effects scattered across the genome. Under this model, specific chromosomal regions are not expected to explain a significant amount of genetic variation. The polygenic model provides the log-likelihood against which to test the alternative hypothesis of linkage. In addition, it yields information about the relative influence of the different variance components on the total variation.

Fixed effects can be included to account for known influences on the phenotypic mean, while the remaining variance is partitioned among specified random effects (Lynch and Walsh 1998; Williams and Blangero 1999). In the simplest case, the random effects will include just the additive genetic value such that:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

Where **y** is a vector of records on individuals; β is a vector of fixed effects, **a** is a vector of additive genetic effects (or breeding values) estimated on the basis of the coefficient of co-ancestry between any pair of individuals in the pedigree; **e** is a vector of residual effects. **X** and **Z** are design matrices relating records to the appropriate fixed or random effects. The additive genetic relationship matrix created from the pedigree file incorporated information from all known and inferred relatives, of both sexes, correctly weighted for relatedness. Where different measurements of the same trait, on the same individual, were available at different life stages, the permanent environmental effect grouped the repeated measurements to determine the environmental variance between individuals that arose from long-term or non-localized conditions.

Heritability (h^2), permanent environment effect (c^2), and residual effect (e^2) were calculated as the ratio of the relative variance component (V_A, additive genetic variance;

V_C, permanent environmental variance; V_E, residual variance) to total phenotypic variance (V_P), i.e. $h^2 = V_A/V_P$; $c^2 = V_C/V_P$; $e^2 = V_E/V_P$.

The coefficient of variation, expressed as a percentage, (CV) standardizes the variance by the trait mean instead of the total variance, and it is calculated as the ratio of the standard deviation (square root of the variance) to the mean times 100. Therefore:

$$CV_i = 100V_i^{1/2} / \overline{x},$$

where the subscript i stands for the additive genetic (A), permanent environment (C), and residual components (E) and $\overline{x}$ is the trait mean. As the coefficient of variation standardizes the variance by the trait mean instead of by the total phenotypic variance, it provides an alternative to the heritability to compare the degree of dispersion of the measures for traits with different mean values.

Variance components were estimated by the restricted maximum likelihood procedure (Lynch and Walsh 1998) implemented in the software package ASReml (Gilmour et al. 2002). The estimated parameters of the variance components under the null hypothesis of no QTL effects are listed in Table 4.1.

QTL mapping – To map putative segregating QTL, an IBD (identity by descent) matrix estimated at any given map position was fitted in the polygenic model described above as an additional random effect (George et al. 2000):

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

Where **q** is a vector of additive QTL effect. As the IBD matrix represents the allele sharing probability between any pair of individuals in a specific genomic region, the significance of its effect on the study trait is evidence of segregating QTL in the region

being tested. IBD sharing statistics were estimated using pedigree relationships, marker data, and map distances described above and in Beraldi et al. (2006). For an initial scan, IBD matrices and variance components were estimated every 10 cM. Putative QTL regions, i.e. those reaching a LOD score of at least 1, were then scanned every 1 cM. The IBD sharing analysis was performed by a Markov chain Monte Carlo (MCMC) procedure which is based on a stochastic process (gene-drop simulations) and as such does not provide an exact result, but allows the handling of very large and complex pedigrees. After a burn-in period of 1000 cycles, 100,000 MCMC iterations were performed and sample statistics were stored every ten iterations. This process was implemented in the program Loki (Heath 1997). The IBD matrices were then inverted and fitted one by one in ASReml using a program written by one of the authors (AFM) to automate the process of inputting and storing the results. LOD scores were calculated as the difference in log-likelihood between QTL and polygenic model according to the equation:

$$LOD = (L_1 - L_0) / \ln(10)$$

Where L_1 is the natural log-likelihood of the QTL model and L_0 the natural log-likelihood of the polygenic model.

Genome-wide suggestive and significant thresholds were obtained by solving equation 1 of Lander and Kruglyak (1995) assuming a map length of 33.5 Morgans spanning 27 chromosomes. Solutions were 1.9 and 3.3, respectively. The genome-wide significance (LOD= 3.3) corresponds to the probability of finding a false positive every 20 genome scans; the suggestive significance (LOD= 1.9) corresponds to the probability of finding a false positive once per genome scan (Lander and Kruglyak, 1995). Here, all the LOD scores exceeding the arbitrary threshold of 1 are reported. For LOD scores above the suggestive threshold, confidence intervals for the presence of a putative QTL

were defined by the map range within a one-LOD score drop from the peak value; which is equivalent to approximately 95% confidence interval (Lander and Botstein, 1989).

4.4. Results and Discussion

Six (non independent) traits reflecting the resistance of Soay sheep to gastrointestinal strongyles, coccidia, and keds at ages four months and 16 months or older were modelled in order to estimate the genetic and environmental contribution to the variation of the traits and, consequently, to detect major genes (QTL) responsible for the additive genetic variation. QTL mapping by means of variance component analysis requires the estimation of the variance components under the null hypothesis of no segregating QTL. This step offers the opportunity to analyse the quantitative genetics parameters of the study population and, therefore, it is presented and discussed separately from the QTL mapping results.

Variance component analysis – Results of the variance component analysis under the polygenic model are presented in Table 4.1. In lambs, the additive genetic component of strongyle FEC and coccidia FOC accounted for a moderately high proportion of the phenotypic variation although the estimates were not very precise due to the large standard deviations. The heritabilities of strongyle FEC and coccidia FOC in lambs were similar, being 0.26 ± 0.12 and 0.22 ± 0.21, respectively. In adults, no genetic variation was detected for strongyle FEC and very low heritability was detected for coccidia FOC (h^2= 0.06 ± 0.03). The estimates of heritability reported by Coltman et al. (2001a) for strongyles FEC in Soay sheep were between 0.11 and 0.14. The inconsistency between Coltman et al.'s results and the present study could be explained by differences in the pedigree and data selection. In particular, the estimates of Coltman et al (2001a) are based on animals of any age whereas in this study we differentiated between lambs

(four months old animals) and adults (animals older than four months). Also, higher estimates of heritability in this study may be explained by more reliable inference of parentage in this study. The genotyping of more than 200 markers in the genome scan allowed the detection of pedigree errors that can downwardly bias the estimate of genetic parameters (Charmantier and Reale, 2005). No previous studies in Soay sheep have investigated the genetics of coccidia FOC so that it is not possible to make comparisons within the study population. With respect to domestic sheep, the estimates of FEC heritability in Soay sheep reported here do not particularly differ from farmed or experimental populations. However, as the heritability is a property of the population and not of the species, care should be taken in comparing Soay and domestic sheep because of the differences in life history and environment between a free-living population and managed, selected flocks.

With respect to the coefficients of variation, the CV_A of strongyles FEC in lambs was about three times the CV_A of coccidia FOC in lambs (18.08 and 6.32 respectively) although the heritability of the two traits was similar (0.26 and 0.22 respectively). This suggests that there is greater genetic variation responsible for strongyle FEC then for coccidia FOC, but also that the phenotypic variation of strongyle FEC is higher then the phenotypic variation of coccidia FOC. (By definition the CV_A increases with the additive genetic variance and is independent of the phenotypic variance, whereas the heritability increases with the additive genetic variance but decreases with the phenotypic variance; see Materials and Methods).

The adult datasets of strongyle FEC and coccidia FOC and ked count in both lambs and adults showed little or zero heritable variation (Table 4.1). However, the CV_A in adult coccidea FOC (10.66) is higher than the one in lambs (6.32). This suggests that

the genetic variation in adults is overwhelmed by the environmental variation so that the genetic component makes little contribution to phenotype. The same speculation could be applied to strongyle FEC and ked count but in this case the fact that the means and variances are similar and close to zero make the coefficient of variation difficult to interpret due to its mathematical properties. The lack of detectable heritable variation in these datasets could be due to limited sample sizes within the mapping population, and changing environmental conditions masking the additive genetic contribution to the trait variation.

Variance component QTL mapping – Genome scans were performed for strongyle and coccidia FOC in lambs, the two traits with moderate heritability. The LOD score profiles for these traits are shown in Figure 4.1, and the characteristics of the LOD scores higher than one are listed in Table 4.3.

Table 4.3. LOD scores higher than 1 detected for strongyle and coccidia FEC in lambs.

Trait	Dataset	LOD	Chr.	Position (cM)	Flanking markers (cM)[a]	
Strongyles	Lambs	1.58	6	74	BMS360 (4)	McM140 (4)
		1.49	12	44	CSSM3 (1)	MCMA52 (19)
		1.43	1	79	BM6465 (7)	CSAP36E (1)
Coccidia	Lambs	2.68[b]	3	328	CSAP39E (17)	CSSME76 (0)
		2.21[b]	X	3	McM158 (3)	MAF45 (32)
		1.52	3	127	RM96 (6)	BM2818 (14)
		1.13	2	89	CSSM37 (7)	FCB128 (3)

[a] In parentheses the distance (cM) of the flanking markers from the QTL peak. [b] suggestive linkage (LOD>1.9)

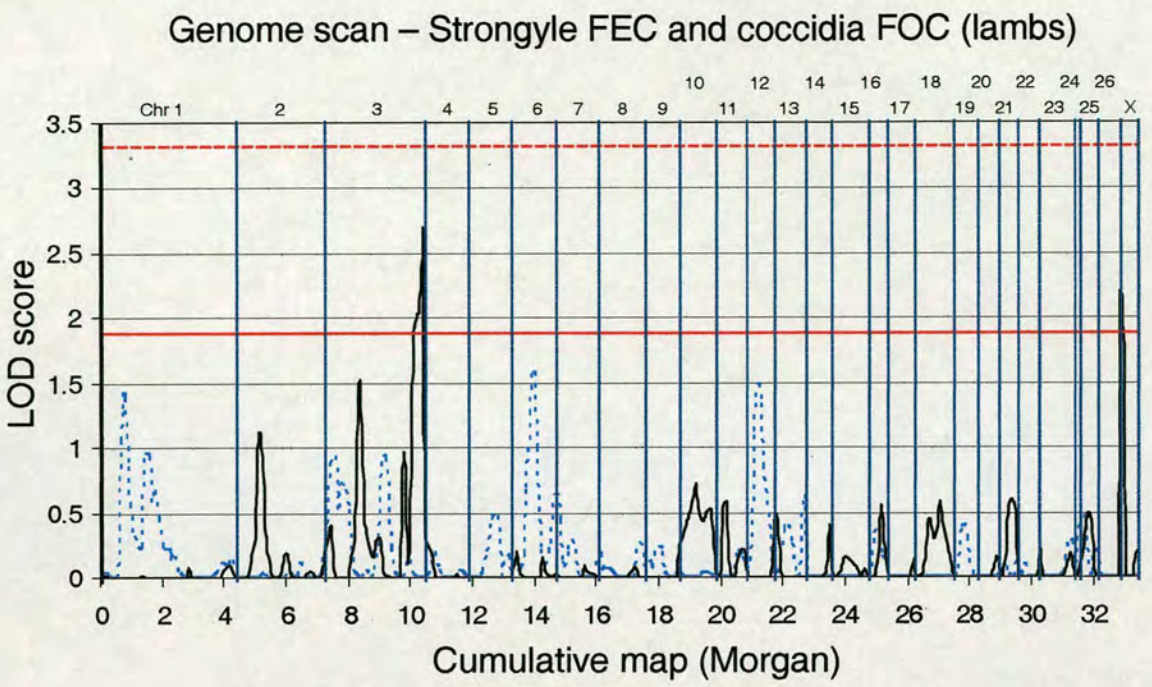

Figure 4.1 – Whole genome scans of strongyle FEC (dashed line) and coccidia FOC (continuous line) in lambs. LOD score values (ordinate) were plotted against genetic position (abscissa, Morgan scale). Dotted horizontal lines show the genome-wide significance threshold (3.3); dashed lines are the suggestive significance threshold (1.9). Vertical lines mark the chromosome boundaries and chromosome names are displayed at the top.

Three LOD scores above 1 but below the suggestive threshold were detected for strongyle FEC (Table 4.3). These were located on chromosome 6 (LOD= 1.58), 12 (LOD= 1.49) and 1 (LOD= 1.43). Beh et al. (2002) detected a suggestive QTL for resistance to *Trichostrongylus colubriformis* in 20 week-old sheep on chromosome 6 and a pointwise significant peak (significant at $p<0.05$ but unadjusted for multiple tests) for 27 week old animals. Beh et al. (2002) also identified one region on each of these chromosomes reaching the pointwise significance. The scan published by Beh et al. (2002) does not report the position of the LOD peaks so that it is not possible to determine whether their peaks correspond to those presented here. To the best of our

knowledge, no study other then Beh's et al. (2002) detected QTL for parasite resistance in the regions reported here.

The scan for coccidia FOC in lambs produced two LOD scores exceeding the suggestive threshold (chromosome 3 with LOD= 2.68 and confidence interval of approximately 30 cM, and chromosome X with LOD= 2.21 and confidence interval of approximately 17 cM; Table 4.3 and Figure 4.2), and two LOD scores exceeding the value of 1 (chromosome 3 with LOD= 1.52 and chromosome 2 with LOD= 1.13; Table 4.3). The LOD score peak for coccidia FEC in lambs on chromosome X is in the vicinity of one of the telomeres. No previous studies have investigated chromosome X for parasite resistance QTL or genes. Davies et al. (2006) identified three regions on chromosome 3 likely to be linked to different traits related to parasite resistance (IgA activity, *Nematodirus* FEC in August, and strongyle FEC in October) and one region on chromosome 2 linked to *Nematodirus* FEC in September. Although the confidence intervals of the present study and of the Davies's et al. (2006) overlap, the statistical significance of the results and the differences in the two approaches make it difficult at this stage to understand whether the two studies identified the same regions. Other studies in domestic sheep did not identified QTL in the regions detected here.

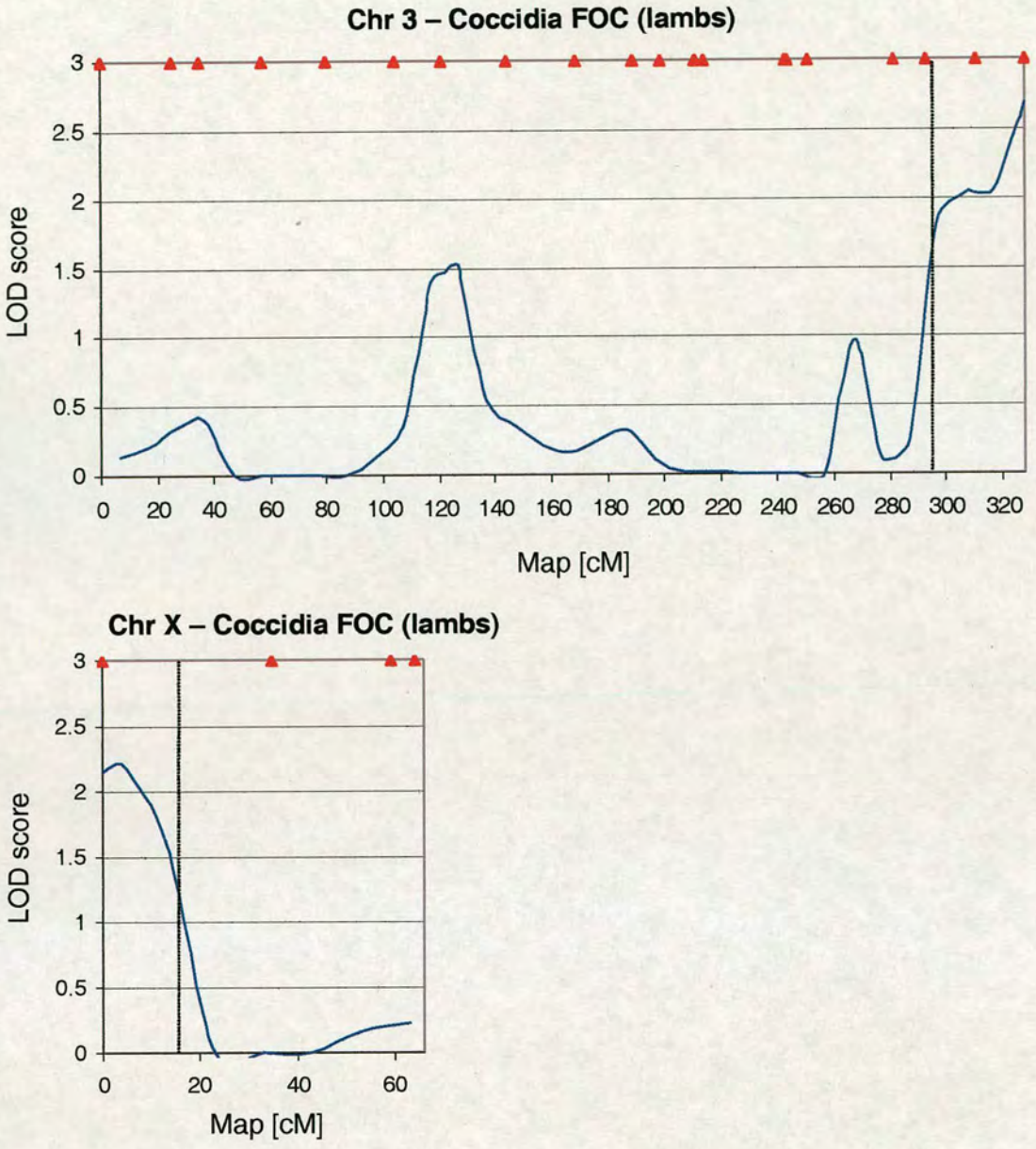

Figure 4.2 – Detailed map positions of the suggestive QTL peaks identified by the genome scan for coccidia FOC in lambs. The abscissas represent the chromosome maps (centimorgan scale). Vertical lines define the 1-LOD drop confidence intervals. Triangles on top of the graphs show the marker positions.

Candidate regions identified in previous association studies in Soay sheep did not produce any relevant evidence of linkage in these genome scans. It is known that the interferon gamma (IFNG) region on chromosome 3 is related to strongyle parasite resistance in domestic sheep (Paterson et al., 1999) and Soay lambs (Coltman et al.,

2001b). The IFNG region, located at approximately 244 cM on the Soay map, did not produce any particular evidence of linkage and it is outside the confidence interval of the suggestive QTL identified for coccidia FOC. The effect of IFNG on strongyle FEC in Soay sheep was estimated by Coltman et al. (2001b) in a dataset larger than the one here analyzed and using general linear model in which the two alleles of a microsatellite in the IFNG gene were fitted as a fixed factor (Coltman et al., 2001b). The association between FEC and microsatellite alleles was significant at p= 0.047 but the variation explained by the microsatellite was very low (0.5% of the total deviance). Consequently, it is not surprising that no QTL was detected in this region considering the differences in datasets and methods of analysis. Similarly, the ADA locus, mapped to chromosome 13 (Beraldi et al., 2006), and the MHC region on chromosome 20 did not produce any evidence of linkage. As for IFNG, the association between the MHC region and strongyle FEC was detected using a larger dataset with a generalized linear model. Failure to detect linkage could be due to insufficient power of the current sample size or lack of informative markers in the target regions although good marker coverage was achieved in the putative regions (Beraldi et al. 2006). However, association studies are prone to produce false positive results because of, for example, population structure (Cardon and Bell, 2001). Population stratification occurs if the sample consists of a number of divergent populations which differ in both candidate-locus frequencies and phenotypic frequency. In this case, an association can be detected in the absence of linkage. It should be noted in this respect that, despite its small size, the Soay sheep population is structured into at least three sub-units genetically and spatially differentiated (Coltman et al., 2003; Charbonnel and Pemberton, 2005).

Conclusions and future developments – This paper describes a genome scan in the free-living Soay sheep to detect QTL associated with resistance to different kinds of

parasites, here measured as strongyle and coccidia FOC and ked count, through variance component analysis. This is one of the first genome-wide scans for parasite resistance performed in sheep and the first which makes use of a free-living population. Future projects aimed at undertaking the same approach may find the methods and findings of this study to be a useful reference; for this reason, results that are liable to be false positives have been presented. Overall, the methods and data described here show that FEC of strongyle and coccidia have moderate heritability in lambs and lower or undetectable heritability in adults. Two genomic regions reached the suggestive linkage threshold which could be confirmed or rejected by the genotyping of additional markers mapping in the target region or by analysing more families.

The power of analysis of the presented study is probably not high enough to detect genes with small effect. Parasite resistance measured as FEC or FOC is the result of a large number of physiological pathways that from the original infection lead to the egg or oocyst count. The target phenotype is therefore a composite of different traits with a strong environmental component. The use of more specific measures of parasite resistance, for example the egg count of individual parasitic species, should make the analysis more powerful as it would focus on a phenotype which would be less affected by the environmental conditions. The identification of single species of nematodes has been initiated by Wimmer et al (2004) using analysis of molecular genetic variation. In order to reduce the statistical noise due to the environment it would be advisable to collect as many measurements as possible in different years on the same individual to detect the component of phenotypic variation due to the permanent environmental conditions.

In the long term, knowledge of the map position of a gene affecting parasite resistance can be used to examine selection at, and molecular evolution of, resistance

genes. The future statistical and biological models aimed at explaining the maintenance of genetic variation in Soay sheep will be improved by accounting for the variability at specific genomic regions known to affect the level of parasitic infection.

Acknowledgments

We thank JG Pilkington, BH Craig and many previous volunteers and project members for collecting field data and genetic samples. We thank A. Wilson for statistical advice. We thank the National Trust for Scotland for granting permission to work on St. Kilda, and QinetiQ for logistical support. The long term data collection on St. Kilda has been supported by Natural Environment Research Council (NERC) and Wellcome Trust grants to T.H. Clutton-Brock, B.T. Grenfell, L.E.B Kruuk, M.J. Crawley and JMP. This study was funded by NERC through its Environmental Genomics thematic programme (grant number NER/T/S/2002/00189).

Chapter 5.

Discussion and future work

This project describes one of the first developments of a linkage map (Figure 2.3) and genome wide scans for QTL in a free-living animal population, thus opening the way to the dissection of the genetic architecture of complex traits in the wild. QTL mapping gives the possibility to investigate the nature of additive genetic variation (Lynch and Walsh 1998) and its maintenance under selection (Fisher 1958, Charlesworth 1987, Barton and Turelli 1989) in order to resolve whether major genes affect the genetic variation of a trait (Mitchell-Olds 1995).

Synthesis of results – A linkage map was constructed by genotyping 247 microsatellite and four allozyme markers on a multigenerational pedigree comprising 588 individuals (Chapter 2). Approximately 90% of the sheep genome was covered (~3350 cM) with an average inter-marker spacing of 15 cM (Figure 2.3 and Table 2.2). Three discrete traits of evolutionary importance (coat colour, coat pattern, and horn type) were mapped via linkage analysis (Figures 2.4 and 2.5). The *Coat colour* locus mapped to chromosome 2; *Coat pattern* mapped to chromosome 13, close to the candidate locus *Agouti*. Finally, *Horn type* mapped to chromosome 10, a location similar to that previously identified in domestic sheep (Montgomery et al., 1996). These findings represent an advance in the dissection of genetic diversity in the wild and provide the foundation for the subsequent QTL analyses.

The linkage map and related mapping panel were used to detect QTL responsible for a variety of morphometric and early developmental traits related to the individual fitness (Chapter 3). The study traits included birth date and weight, considered both as

maternal and offspring traits, fore and hind leg length, body weight, jaw length, and metacarpal length (Table 3.1). Genetic and environmental components of phenotypic variance were estimated for each trait (Table 3.1) and, for those traits showing non-zero heritability, a QTL search was conducted by variance component analysis (Figure 3.1 and Table 3.3). Support for a QTL at genome-wide significance (LOD> 3.3) was found on chromosome 11 for jaw length; suggestive QTL (LOD> 1.9) were found on chromosomes 2 (for birth date as a trait of the lamb), 8 (birth weight as a trait of the lamb) and 15 (adult hindleg length) (Table 3.3 and Figure 3.3).

An approach similar to the one applied for the morphometric traits was used to detect QTL affecting resistance to gastrointestinal parasites and ectoparasitic keds (*Melophagus ovinus*) (Chapter 4). The traits here investigated were the strongyle faecal egg count, the coccidia oocyst count, and the count of keds (Table 4.1). Two genomic regions reached the level of suggestive linkage for coccidia oocyst count in lambs (chromosome 3 and X, respectively; Figures 4.1 and 4.2, Table 4.3).

Prospects to the understanding of genetic variation in the wild – How can the results presented here help in understanding natural genetic variation? Coat colour and horn type are traits under selection in Soay sheep, since dark sheep are heavier then light ones and are positively selected during some high mortality winters (see Discussion in Chapter 2 for references and more information). No convincing hypothesis has been provided so far to explain how selection on coat colour operates and how the light colour morph is maintained in the population. The mapping of the *Coat colour* locus offers the prospect of understanding these issues much better in future. Hypotheses and future work to explain the difference in survival will take advantage of the map position and molecular characterization of the *Coat colour* gene. For example, selection on coat colour can now be analyzed separately for homozygote and heterozygote dark sheep to

determine whether heterozygotes are fitter. Also, an analysis of LD surrounding the *Coat colour* locus could provide information about the origin and evolutionary consequences of coat colour variation. Although the mapping of *Coat pattern* was successful, limited insights can be gained from this locus in terms of understanding natural genetic variation since at the moment no selection acting on this locus has been detected.

The *Horn type* locus was mapped but additional work is required to refine the map position of the responsible gene(s). In Soay sheep, as well as in other ruminants, horn type is involved in mate choice. Previous analyses of Soay sheep have suggested that normal-horned males and scurred females have highest annual breeding success (Clutton-Brock et al., 1997; Stevenson et al., 2004), but that in winters characterized by high mortality, the scurred phenotype is generally favoured in both sexes (Moorcroft et al., 1996). Exactly how these forces maintain variation in the population is the subject of current research and would clearly be helped by being able to distinguish individuals by genotype rather than phenotype. It should be noted, however, that traits encoded by single genes make it difficult to generalize the findings to complex traits responsible of microevolutionary changes.

In order to better understand natural genetic variation, traits related to fitness and presenting continuous phenotypic distributions were investigated by variance component QTL mapping. The traits investigated (measures of early development, body size, and parasite resistance) are quantitative and, in contrast to the simple Mendelian traits described above, are probably more responsible for microevolutionary change. Most of the traits analyzed here have documented relationships with total fitness (see Introductions to Chapter 3 and 4). In general, the linear traits (leg lengths, metacarpal and jaw length) had high values of h^2 and low values of CV_A whereas the body weight traits had lower h^2 but higher CV_A (Table 3.1). For example, although the heritability of

adult hindleg length was found to be more than twice that of adult body weight, the coefficient of variation was higher for body weight than for hindleg length (CV_A= 6.57 vs. 2.80). This trend confirms the intuition that body weight is composed of several underlying traits which confer a higher overall genetic contribution to the trait, but also a higher residual variation which in turn results in a lower ratio V_A/V_P (i.e. heritability).

With respect to the measures of parasite resistance, in lambs the additive genetic component of strongyle FEC and coccidia FOC accounted for a moderately high proportion of the phenotypic variation (Table 4.1) and their heritabilities in lambs were similar (0.26 ± 0.12 and 0.22 ± 0.21, respectively). In adults, no genetic variation was detected for strongyle FEC and very low heritability was detected for coccidia FOC. Compared to domestic sheep, the estimates of FEC heritability in Soay sheep reported here are similar to those from farmed or experimental populations. With respect to the coefficients of variation, the CV_A of strongyle FEC in lambs was about three times the CV_A of coccidia FOC in lambs (18.08 and 6.32 respectively) although the heritability of the two traits was similar (0.26 and 0.22 respectively). This suggests that there is greater genetic variation responsible for strongyle FEC then for coccidia FOC, but also that the phenotypic variation of strongyle FEC is higher than the phenotypic variation of coccidia FOC.

In Tables 3.3 and 4.3 we report estimated effect sizes in order to provide a comparison with similar studies; however, it should be noted that these estimates are inevitably inflated. In fact, the estimation of the QTL heritability and effect from a genome wide scan aimed at mapping is upwardly biased and probably unrealistic, especially when the QTL has weak effect (Goring et al., 2001). This is because the maximum likelihood procedure provides the highest statistical evidence of linkage (LOD

score) where the parameters (QTL effect) are maximized. Therefore, independent datasets should be analysed if both QTL position and effect are desired (Goring et al., 2001). No attempt has been made in this thesis to quantify the pressure of selection on the candidate QTL. Also, no higher order analyses such as multivariate or random regression analysis, which could investigate the change in additive genetic variance with the life stage of the animals, were performed. Currently, the significance of the results was too low to test more complex models.

Limitations of the project – Overall, the results presented here constitute a starting point to understanding the nature of quantitative genetic variation in Soay sheep, and ideally in other wild populations. However, more work is required before this task can be accomplished successfully and some limits to the project should be noted. The candidate QTL need to be confirmed and refined (see below) and, at the moment, the possibility that these QTL are false positives cannot be rejected. In addition, the mapping panel used in this study may not be representative of the Soay sheep population since, to increase the power of QTL detection, it was selected to include the largest families. Therefore, the consistency of the findings in a larger, random sample of pedigrees of the Soay sheep population should, ideally, be demonstrated. QTL mapping typically requires large sample sizes and good data quality (Flint and Mott 2001). A large dataset is particularly required for traits with low heritability and high year to year variation (e.g. body size or parasite resistance). Furthermore, traits with high evolutionary importance (e.g. body weight, survival, number of offspring) are expected to show a low heritability. In choosing a target trait, therefore, a trade-off has to be made between maximizing power of analysis and evolutionary interest. Because of the time period over which the data were collected, most of the animals belonging to the mapping pedigree are now dead and, for any living animal, there is no guarantee that an individual can be re-

captured and confidently tracked from one year to another. These limits will reduce the capability to detect effects of interest like the genotype by environment interaction or the variation of a trait across different ontogenetic stages. QTL mapping in natural populations is becoming more widespread and these limits will be probably shared by most studies (Slate 2005). Therefore, it is foreseeable that the methods applied and the shortcomings encountered here will provide a guideline for future similar projects.

Future developments - With respect to the discrete traits, follow-up studies on the *Coat colour* locus have identified the responsible gene (TYRP1) and the putative causal mutation (Gratten et al. In press). The *Horn type* region will be further examined to refine the map position of the target locus and to identify putative responsible genes. In addition, extensive linkage disequilibrium surrounds the *Coat colour* locus (our unpublished results). Understanding the origin and pattern of this linkage disequilibrium and its relationship with *Coat colour* could reveal insights into events of selection and/or population introgression occurring in this region. In general, the Soay linkage map can provide a backbone to investigate patterns of linkage disequilibrium on a larger scale in the genome.

A number of different strategies can lead to the improvement of the QTL analyses to confirm or refine the results here shown. The genotyping of additional markers in the critical regions will provide a denser marker map. This, in turn, will give more accurate estimates of the IBD coefficients to generate more reliable tests for the presence of a QTL. This strategy will be particularly advantageous when the markers in target regions are sparse or have low information content. Additional families can also be typed to increase the sample size. The choice of the families will be based on their number of offspring as larger families provide higher statistical power. The genotyping of an

independent dataset could also provide a better estimate of the QTL effect without the inflation induced by the mapping procedure (Goring et al., 2001). Alternatively, families could be chosen from the tails of the breeding value distribution for the study trait(s) in order to contrast marker genotypes (IBD) from individuals with presumably discordant QTL genotypes. With respect to the statistical procedure, the power and accuracy of QTL detection can be increased by analysing different traits simultaneously and/or by making use of a combination of linkage and linkage disequilibrium (Meuwissen and Goddard, 2004; Hernandez-Sanchez et al., 2006). The simultaneous analysis of different traits is based on the assumption that a QTL has a pleiotropic effect on the traits included so that information from different datasets can be combined to achieve higher accuracy of analysis. This approach, however, will only be effective when the genetic correlation between traits is high. In addition, the number of parameters to be estimated increases rapidly with the number of traits studied so that the method could be inefficient in small datasets. In a variance component framework, linkage disequilibrium can be modelled by allowing the founder QTL alleles to be related. IBD coefficients between founder QTL will be estimated using the haplotypes from surrounding markers. Linkage disequilibrium mapping alone, which makes use of historical recombinations, can lead to false positive results as linkage disequilibrium can extend over large distances. Linkage mapping, however, will not confirm such associations. The combination of linkage and linkage disequilibrium is expected to give more robust and accurate estimates of QTL positions (Meuwissen and Goddard, 2000, 2001; Meuwissen et al., 2002; Meuwissen and Goddard, 2004).

Once a QTL has been confirmed, its biological relevance and role can be formally tested in the same variance components procedure. By means of random regression analysis (Macgregor et al., 2005), interactions between QTL and environment or between

QTL and age can be investigated to detect when and how a QTL is expressed. In addition, the breeding values associated with a QTL could be monitored across generations to detect selection acting on the target regions. Although the size of the dataset may become a limiting factor, these approaches may be worth pursuing as they would offer an unprecedented description of the relationship between genetics and environment.

In the long term, alternative strategies may prove to be more effective than linkage mapping to locate and identify the genomic regions responsible for a trait of interest. Technical advances in SNP genotyping and genome sequencing may allow relatively inexpensive screening of many markers (in the range of hundreds of thousands) scattered across the genome and the sequence analysis of virtually any gene. In this scenario linkage disequilibrium mapping will provide more powerful analysis than linkage mapping without the obligatory need for pedigree data. Besides, a very dense SNP map may reveal signatures of selection and genome-wide features (e.g. distribution of recombination hot-spots) by highlighting patterns of linkage disequilibrium and genetic variation without the need for phenotypic data. Genome sequencing projects are underway in sheep (http://www.livestockgenomics.csiro.au/sheep.shtml) and cattle (http://www.hgsc.bcm.tmc.edu/projects/bovine/). These resources will help in identifying candidate genes responsible for traits of interest and will facilitate comparative mapping.

The genetic basis of trait variation does not only include allelic variation but also, and possibly more important, differences in gene expression and epigenetic changes such as DNA methylation and chromatin structure. Epigenetic changes are heritable but reversible and as such can control gene expression in a stable but flexible way. A

thorough screening of these sources of variation will complement the identification of QTL regions. The mapping of QTL is a step towards the understanding of quantitative genetic variation, but an exhaustive explanation will require a fuller view of the mechanisms leading from genotype to phenotype.

References

Abecasis, G.R., Ghosh, D., Nichols, T.E., 2005. Linkage disequilibrium: ancient history drives the new genetics. Hum Hered 59, 118-124.

Abiola, O., Angel, J.M., Avner, P., Bachmanov, A.A., Belknap, J.K., Bennett, B., Blankenhorn, E.P., Blizard, D.A., Bolivar, V., Brockmann, G.A., Buck, K.J., Bureau, J.F., Casley, W.L., Chesler, E.J., Cheverud, J.M., Churchill, G.A., Cook, M., Crabbe, J.C., Crusio, W.E., Darvasi, A., de Haan, G., Dermant, P., Doerge, R.W., Elliot, R.W., Farber, C.R., Flaherty, L., Flint, J., Gershenfeld, H., Gibson, J.P., Gu, J., Gu, W., Himmelbauer, H., Hitzemann, R., Hsu, H.C., Hunter, K., Iraqi, F.F., Jansen, R.C., Johnson, T.E., Jones, B.C., Kempermann, G., Lammert, F., Lu, L., Manly, K.F., Matthews, D.B., Medrano, J.F., Mehrabian, M., Mittlemann, G., Mock, B.A., Mogil, J.S., Montagutelli, X., Morahan, G., Mountz, J.D., Nagase, H., Nowakowski, R.S., O'Hara, B.F., Osadchuk, A.V., Paigen, B., Palmer, A.A., Peirce, J.L., Pomp, D., Rosemann, M., Rosen, G.D., Schalkwyk, L.C., Seltzer, Z., Settle, S., Shimomura, K., Shou, S., Sikela, J.M., Siracusa, L.D., Spearow, J.L., Teuscher, C., Threadgill, D.W., Toth, L.A., Toye, A.A., Vadasz, C., Van Zant, G., Wakeland, E., Williams, R.W., Zhang, H.G., Zou, F., 2003. The nature and identification of quantitative trait loci: a community's view. Nat Rev Genet 4, 911-916.

Allison, D.B., Neale, M.C., Zannolli, R., Schork, N.J., Amos, C.I., Blangero, J., 1999. Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. Am J Hum Genet 65, 531-544.

Almasy, L., Blangero, J., 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62, 1198-1211.

Amos, C.I., 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54, 535-543.

Andersson, L., 2001. Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet 2, 130-138.

Baker, R.L., Watson, T.G., Bisset, S.A., Vlassoff, A., Douch, P.G.C., 1991. Breeding sheep in New Zealand for resistance to internal parasites: research results and commercial application.. In: Gray, G.D.W., R. R (Ed.), Breeding for disease resistance in sheep, Australian Wool Corporation, Melbourne, pp. 19-32.

Bamshad, M., Wooding, S.P., 2003. Signatures of natural selection in the human genome. Nat Rev Genet 4, 99-111.

Barton, N.H., Keightley, P.D., 2002. Understanding quantitative genetic variation. Nat Rev Genet 3, 11-21.

Bennett, D.C., Lamoreux, M.L., 2003. The color loci of mice - a genetic century. Pigment Cell Res 16, 333-344.

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M., Pemberton, J.M., 2006. Development of a Linkage Map and Mapping of Phenotypic Polymorphisms in a Free-Living Population of Soay Sheep (Ovis aries). Genetics 173, 1521-1537.

Beraldi, D., McRae, A.F., Gratten, J., Slate, J., Visscher, P.M., Pemberton, J.M. Quantitative trait loci (QTL) mapping of resistance to strongyle and coccidia in the free-living Soay sheep (*Ovis aries*). Int J Parasitol. In press.

Bishop, S.C., Bairden, K., McKellar, Q.A., Park, M., Stear, M.J., 1996. Genetic parameters for faecal egg count following mixed, natural, predominantly Ostertagia circumcincta infection and relationships with live weight in young lambs. Animal Science 63, 423-428.

Bishop, S.C., Stear, M.J., 2003. Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections. Vet Parasitol 115, 147-166.

Botstein, D., Risch, N., 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl, 228-237.

Boyd, J.M., Doney, J.M., Gunn, R.G., 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, 129-143.

Brem, R.B., Kruglyak, L., 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102, 1572-1577.

Campbell, R.N., 1974. St. Kilda and its Sheep. Island Survivors: the Ecology of the Soay Sheep of St. Kilda. The Athlone Press, University of London.

Cardon, L.R., Bell, J.I., 2001. Association study designs for complex diseases. Nat Rev Genet 2, 91-99.

Carlborg, O., Haley, C.S., 2004. Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5, 618-625.

Carlborg, O., Jacobsson, L., Åhgren, P., Siegel, P., Andersson, L., 2006. Epistasis and the release of genetic variation during long-term selection. Nat Genet 38, 418-420.

Casas, E., Shackelford, S.D., Keele, J.W., Koohmaraie, M., Smith, T.P., Stone, R.T., 2003. Detection of quantitative trait loci for growth and carcass composition in cattle. J Anim Sci 81, 2976-2983.

Casas, E., Keele, J.W., Shackelford, S.D., Koohmaraie, M., Stone, R.T., 2004. Identification of quantitative trait loci for growth and carcass composition in cattle. Anim Genet 35, 2-6.

Charbonnel, N., Pemberton, J., 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. 95, 377.

Charmantier, A., Reale, D., 2005. How do misassigned paternities affect the estimation of heritability in the wild? Molecular Ecology 14, 2839.

Clutton-Brock, T., Price, O., Albon, S., Jewell, P., 1992. Early development and population fluctuations in Soay sheep. J. Anim. Ecol. 61, 381-396.

Clutton-Brock, T.H., Wilson, K., Stevenson, I.R., 1997. Density-dependent selection on horn phenotype in Soay sheep. Philos Trans R Soc Lond B Biol Sci 352, 839-850.

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

Clutton-Brock, T.H., Pemberton, J.M., 2004. Soay Sheep Dynamics and Selection in an Island Population. Cambridge University Press, Cambridge.

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

Coltman, D.W., Smith, J.A., Bancroft, D.R., Pilkington, J., MacColl, A.D., Clutton-Brock, T.H., Pemberton, J.M., 1999. Density-Dependent Variation in Lifetime Breeding Success and Natural and Sexual Selection in Soay Rams. Am Nat 154, 730-746.

Coltman, D.W., Pilkington, J., Kruuk, L.E., Wilson, K., Pemberton, J.M., 2001a. Positive genetic correlation between parasite resistance and body size in a free-living ungulate population. Evolution Int J Org Evolution 55, 2116-2125.

Coltman, D.W., Wilson, K., Pilkington, J.G., Stear, M.J., Pemberton, J.M., 2001b. A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. Parasitology 122, 571-582.

Coltman, D.W., Festa-Bianchet, M., Jorgenson, J.T., Strobeck, C., 2002. Age-dependent sexual selection in bighorn rams. Proc Biol Sci 269, 165-172.

Coltman, D.W., Pilkington, J.G., Pemberton, J.M., 2003. Fine-scale genetic structure in a free-living ungulate population. Mol Ecol 12, 733-742.

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

Coltman, D.W., 2005. Testing marker-based estimates of heritability in the wild. Mol Ecol 14, 2593-2599.

Conner, J.K., 2002. Genetic mechanisms of floral trait correlations in a natural population. Nature 420, 407-410.

Coulson, T., Albon, S., Pilkington, J., 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., Pemberton, J.M., Clutton-Brock, T.H., Crawley, M.J., Grenfell, B.T., 2001. Age, sex, density, winter weather, and population crashes in Soay sheep. Science 292, 1528-1531.

Craig, B.H., Pilkington, J.G., Kruuk, L.E.B., Pemberton, J.M., 2006a. Epidemiology of parasitic protozoan infections in Soay sheep (Ovis aries) on St. Kilda. Parasitology In press.

Craig, B.H., Pilkington, J.G., Pemberton, J.M., 2006b. Gastrointestinal nematode species burdens and host mortality in a feral sheep population. Parasitology In press.

Crawford, A.M., Paterson, K.A., Dodds, K.G., Diez-Tascon, C., Williamson, P.A., Thomsom, M.R., Bisset, S.A., Beattie, A.E., Greer, G.J., Green, R.S., Wheeler, R., Shaw, R.J., Knowler, K., McEwan, J.C., 2006. Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. BMC Genomics 7.

Csilléry, K., Johnson, T., Beraldi, D., Clutton-Brock, T.H., Coltman, D., Hansson, B., Spong, G., Pemberton, J.M., 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics In press.

Davies, G., Stear, M.J., Benothman, M., Abuagob, O., Kerr, A., Mitchell, S., Bishop, S.C., 2006. Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity.

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, 353-361.

Dominik, S., 2005. Quantitative trait loci for internal nematode resistance in sheep: a review. Genet Sel Evol 37 Suppl 1, S83-96.

Doney, J.M., Ryder, M.L., Gunn, R.G., Grubb, P., 1974. Colour, Conformation, Affinities, Fleece and Patterns of Inheritance of the Soay Sheep. Islands Survivors: The Ecology of the Soay sheep of St. Kilda. The Athlone Press, University of London.

Dykhuizen, D.E., 1998. Santa Rosalia revisited: Why are there so many species of bacteria? Antonie van Leeuwenhoek 73, 25.

Elston, R.C., Stewart, J., 1973. The analysis of quantitative traits for simple genetic models from parental, F 1 and backcross data. Genetics 73, 695-711.

Endler, J.A., 1986. Distribution of Selection Coefficients and Differentials in Natural Populations, Natural Selection in the Wild, Monographs in Population Biology, Princeton University Press, Princeton, pp. 203-223.

Erickson, D., 2005. Quantitative trait loci: Mapping the future of QTL's. Heredity 95, 417-418.

Erickson, D.L., Fenster, C.B., Stenoien, H.K., Price, D., 2004. Quantitative trait locus analyses and the study of evolutionary process. Mol Ecol 13, 2505-2522.

Evans, G.J., Giuffra, E., Sanchez, A., Kerje, S., Davalos, G., Vidal, O., Illan, S., Noguera, J.L., Varona, L., Velander, I., Southwood, O.I., de Koning, D.J., Haley, C.S., Plastow, G.S., Andersson, L., 2003. Identification of quantitative trait loci for production traits in commercial pig populations. Genetics 164, 621-627.

Falconer, D.S., 1989. Introduction to Quantitative Genetics. Longman, New York.

Fisher, R.A., 1918. The correlation between relatives on the supposition of mendelian inheritance. Trans. of the Roy. Soc. of Edinburgh 52, 399-433.

Fisher, R.A., 1958. The Genetical Theory of Natural Selection. Dover Publications, New York.

Flint, J., Mott, R., 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. Nat Rev Genet 2, 437-445.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., Tanksley, S.D., 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289, 85-88.

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

Gill, H.S., Altmann, K., Cross, M.L., Husband, A.J., 2000. Induction of T helper 1- and T helper 2-type immune responses duringHaemonchus contortus infection in sheep. Immunology 99, 458.

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

Glazier, A.M., Nadeau, J.H., Aitman, T.J., 2002. Finding genes that underlie complex traits. Science 298, 2345-2349.

Goring, H.H., Terwilliger, J.D., Blangero, J., 2001. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69, 1357-1369.

Gratten J., Beraldi D., Lowder B.V., McRae A.F., Visscher P.M., Pemberton P.M., Slate J. A single nucleotide substitution in TYRP1 is responsible for dark-light coat colour variation in a free-living population of Soay sheep. Proc R Soc Lond B Biol Sci. In press.

Green, P., Falls, K., Crooks, S., 1990. Documentation for CRI-MAP. Washington University, St Louis.

Gulland, F.M., Fox, M., 1992. Epidemiology of nematode infections of Soay sheep (Ovis aries L.) on St Kilda. Parasitology 105 (Pt 3), 481-492.

Gulland, F.M., Albon, S.D., Pemberton, J.M., Moorcroft, P.R., Clutton-Brock, T.H., 1993. Parasite-associated polymorphism in a cyclic ungulate population. Proc R Soc Lond B Biol Sci 254, 7-13.

Haldane, J.B.S., 1949. Disease and evolution. . Ric. Scientifica 19, 68-76.

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

Hansson, B., Akesson, M., Slate, J., Pemberton, J.M., 2005. Linkage mapping reveals sex-dimorphic map distances in a passerine bird. Proc Biol Sci 272, 2289-2298.

Hawthorne, D.J., Via, S., 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412, 904-907.

Heath, S.C., 1997. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61, 748-760.

Hernandez-Sanchez, J., Haley, C.S., Woolliams, J.A., 2006. Prediction of IBD based on population history for fine gene mapping. Genet Sel Evol 38, 231-252.

Hill, W.G., 1982. Rates of change in quantitative traits from fixation of new mutations. Proc Natl Acad Sci U S A 79, 142-145.

Hoeschele, I., Uimari, P., Grignola, F.E., Zhang, Q., Gage, K.M., 1997. Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics 147, 1445-1457.

Houle, D., 1992. Comparing evolvability and variability of quantitative traits. Genetics 130, 195-204.

Houle, D., Morikawa, B., Lynch, M., 1996. Comparing mutational variabilities. Genetics 143, 1467-1483.

Hudson, P.J., Dobson, A.P., Newborn, D., 1998. Prevention of population cycles by parasite removal. Science 282, 2256-2258.

Huyse, T., Poulin, R., Theron, A., 2005. Speciation in parasites: a population genetics approach. Trends Parasitol 21, 469-475.

Johnson, P.L., McEwan, J.C., Dodds, K.G., Purchas, R.W., Blair, H.T., 2005. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. J Anim Sci 83, 1988-2000.

Jones, O.R., Crawley, M.J., Pilkington, J.G., Pemberton, J.M., 2005. Predictors of early survival in Soay sheep: cohort-, maternal- and individual-level variation. Proc Biol Sci 272, 2619-2625.

Kaplan, R.M., 2004. Drug resistance in nematodes of veterinary importance: a status report. Trends Parasitol 20, 477-481.

Khatkar, M.S., Thomson, P.C., Tammen, I., Raadsma, H.W., 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet Sel Evol 36, 163-190.

Kondrashov, A.S., Turelli, M., 1992. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. Genetics 132, 603-618.

Kroymann, J., Mitchell-Olds, T., 2005. Epistasis and balanced polymorphism influencing complex trait variation. Nature 435, 95-98.

Kruuk, L.E., Clutton-Brock, T.H., Rose, K.E., Guinness, F.E., 1999. Early determinants of lifetime reproductive success differ between the sexes in red deer. Proc Biol Sci 266, 1655-1661.

Kruuk, L.E., Clutton-Brock, T.H., Slate, J., Pemberton, J.M., Brotherstone, S., Guinness, F.E., 2000. Heritability of fitness in a wild mammal population. Proc Natl Acad Sci U S A 97, 698-703.

Kruuk, L.E., 2004. Estimating genetic parameters in natural populations using the "animal model". Philos Trans R Soc Lond B Biol Sci 359, 873-890.

Lander, E., Kruglyak, L., 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11, 241-247.

Lander, E.S., Botstein, D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185-199.

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R., Dudley, J.W., 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168, 2141-2155.

Lexer, C., Welch, M.E., Durphy, J.L., Rieseberg, L.H., 2003. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of Helianthus paradoxus, a diploid hybrid species. Mol Ecol 12, 1225-1235.

Lymbery, A.J., 1996. Finding genetic markers for complex phenotypic traits in parasites. Int J Parasitol 26, 7-17.

Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates.

Macgregor, S., Knott, S.A., White, I., Visscher, P.M., 2005. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. Genetics 171, 1365-1376.

Mackay, T.F., 2001. Quantitative trait loci in Drosophila. Nat Rev Genet 2, 11-20.

Mackay, T.F., 2004. The genetic architecture of quantitative traits: lessons from Drosophila. Curr Opin Genet Dev 14, 253-257.

Maddox, J., 2003. Australian Sheep Gene Mapping Web Site. http://rubens.its.unimelb.edu.au/~jillm/jill.htm.

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., van Hest, B., 2001. An enhanced linkage map of the sheep genome comprising more than 1000 loci. Genome Res 11, 1275-1289.

MAFF, 1986. Manual of veterinary parasitological laboratory techniques. HMSO, London.

Marshall, T.C., Slate, J., Kruuk, L.E., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7, 639-655.

McCleery, R.H., Pettifor, R.A., Armbruster, P., Meyer, K., Sheldon, B.C., Perrins, C.M., 2004. Components of variance underlying fitness in a natural population of the great tit Parus major. Am Nat 164, E62-73.

McEwan, J.C., Mason, P., Baker, R.L., Clarke, J.N., Hickey, S.M., Turner, K., 1992. Effect of selection for productive traits on internal parasite resistance in sheep. Proc. NZ Soc. Anim. Prod. 52, 53-56.

McRae, A.F., Pemberton, J.M., Visscher, P.M., 2005. Modelling Linkage Disequilibrium in Natural Populations: The Example of the Soay Sheep Population of St. Kilda, Scotland. Genetics 171, 251-258.

McRae, A.F., Beraldi, D., 2006. Examination of a region showing linkage map discrepancies across sheep breeds. Mamm Genome 17, 346-353.

Meuwissen, T.H., 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., Goddard, M.E., 2001. Prediction of identity by descent probabilities from marker-haplotypes. Genet Sel Evol 33, 605-634.

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

Meuwissen, T.H., Goddard, M.E., 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genet Sel Evol 36, 261-279.

Milner, J.M., Albon, S.D., Illius, A.W., Pemberton, J.M., H., C.-B.T., 1999. Repeated selection of morphometric traits in the Soay sheep on St Kilda Journal of Animal Ecology 68, 472-488.

Milner, J.M., Pemberton, J.M., Brotherstone, S., Albon, S.D., 2000. Estimating variance components and heritabilities in the wild: a case study using the `animal model' approach. J . EVOL. BIOL. 13, 804-813.

Milner, J.M., Albon, S.D., Kruuk, L.E.B., Pemberton, J.M., 2004. Selection on phenotype. In: Clutton-Brock, T.H.,Pemberton, J.M. (Eds.), Soay Sheep

Dynamics and Selection in an Island Population, Cambridge University Press, Cambridge, pp. 190-216.

Montgomery, G.W., Henry, H.M., Dodds, K.G., Beattie, A.E., Wuliji, T., Crawford, A.M., 1996. Mapping the Horns (Ho) locus in sheep: a further locus controlling horn development in domestic animals. J Hered 87, 358-363.

Moorcroft, P.R., Albon, S.D., Pemberton, J.M., Stevenson, I.R., Clutton-Brock, T.H., 1996. Density-dependent selection in a fluctuating ungulate population. Proc R Soc Lond B Biol Sci 263, 31-38.

Morton, N.E., 1955. Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7, 277-318.

Mousseau, T.A., Roff, D.A., 1987. Natural selection and the heritability of fitness components. Heredity 59 (Pt 2), 181-197.

Nadler, S.A., 1995. Microevolution and the genetic structure of parasite populations. J Parasitol 81, 395-403.

Nyholt, D.R., 2000. All LODs are not created equal. Am J Hum Genet 67, 282-288.

O'Connell, J.R., Weeks, D.E., 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63, 259-266.

Olendorf, R., Rodd, F.H., Punzalan, D., Houde, A.E., Hurt, C., Reznick, D.N., Hughes, K.A., 2006. Frequency-dependent survival in natural guppy populations. Nature 441, 633.

Olson, S., 2002. Population genetics. Seeking the signs of selection. Science 298, 1324-1325.

Overall, A.D., Byrne, K.A., Pilkington, J.G., Pemberton, J.M., 2005. Heterozygosity, inbreeding and neonatal traits in Soay sheep on St Kilda. Mol Ecol 14, 3383-3393.

Parsons, Y.M., Fleet, M.R., Cooper, D.W., 1999. The Agouti gene: a positional candidate for recessive self-colour pigmentation in the Australian Merino. Australian Journal of Agricultural Research, 1099-1103.

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S., Tanksley, S.D., 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127, 181-197.

Paterson, K.A., McEwan, J.C., Dodds, K.G., Morris, C.A., Crawford, A.M., 1999. Fine mapping a locus affecting host resistance to internal parasites in sheep. Proc Assoc Adv Anim Breed Genet 13, 91-94.

Paterson, S., 1998. Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. J Hered 89, 289-294.

Paterson, S., Wilson, K., Pemberton, J.M., 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. Proc Natl Acad Sci U S A 95, 3714-3719.

Preston, B.T., Stevenson, I.R., Pemberton, J.M., Coltman, D.W., Wilson, K., 2003. Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. Proc R Soc Lond B Biol Sci 270, 633-640.

Prichard, R., 1994. Anthelmintic resistance. Vet Parasitol 54, 259-268.

Risch, N., 2000. Searching for genetic determinants in the new millennium. Nature 405, 847-856.

Roff, D.A., Mousseau, T.A., 1987. Quantitative genetics and fitness: lessons from Drosophila. Heredity 58 (Pt 1), 103-118.

Roff, D.A., Simons, A.M., 1997. The quantitative genetics of wing dimorphism under laboratory and "field" conditions in the cricket Gryllus pennsylvanicus. Heredity 78, 235–240.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832-837.

Schmid, P.E., Tokeshi, M., Schmid-Araya, J.M., 2002. Scaling in stream communities. Proc Biol Sci 269, 2587-2594.

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

Slate, J., 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Molecular Ecology 14, 363-379.

Stear, M.J., Bairden, K., Duncan, J.L., Holmes, P.H., McKellar, Q.A., Park, M., Strain, S., Murray, M., Bishop, S.C., Gettinby, G., 1997. How hosts control worms. Nature 389, 27.

Steel, T., 1988. The Life and Death of St Kilda. Collins, Glasgow.

Stevenson, I.R., Marrow, P., Preston, B.T., Pemberton, J.M., Wilson, K., 2004. Adaptive reproduction strategies. In: Clutton-Brock, T.H.,Pemberton, J.M. (Eds.), Soay sheep Dynamics and selection in an island population, Cambridge University Press, Cambridge, pp. 243-275.

Stewart, J., Elston, R.C., 1973. Biometrical genetics with one or two loci: the inheritance of physiological characters in mice. Genetics 73, 675-693.

Terwilliger, J.D., Ott, J., 1994. Handbook of Human Genetic Linkage. Johns Hopkins University Press, Baltimore.

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

Visscher, P.M., Haley, C.S., Heath, S.C., Muir, W.J., Blackwood, D.H., 1999. Detecting QTLs for uni- and bipolar disorder using a variance component method. Psychiatr Genet 9, 75-84.

Walling, G.A., Visscher, P.M., Wilson, A.D., McTeir, B.L., Simm, G., Bishop, S.C., 2004. Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. J Anim Sci 82, 2234-2245.

Williams, J.T., Blangero, J., 1999. Power of variance component linkage analysis to detect quantitative trait loci. Ann Hum Genet 63, 545-563.

Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D., Byrne, K.A., Kruuk, L.E., 2005a. Maternal genetic effects set the potential for evolution in a free-living vertebrate population. J Evol Biol 18, 405-414.

Wilson, A.J., Pilkington, J.G., Pemberton, J.M., Coltman, D.W., Overall, A.D., Byrne, K.A., Kruuk, L.E., 2005b. Selection on mothers and offspring: whose phenotype is it and does it matter? Evolution Int J Org Evolution 59, 451-463.

Wilson, K., Grenfell, B.T., Pilkington, J.G., Boyd, H.E.G., Gulland, F.M.D., 2004. Parasites and their impact. In: Clutton-Brock, T.H.,Pemberton, J.M. (Eds.), Soay Sheep Dynamics and Selection in an Island Population, Cambridge University Press, Cambridge, pp. 52-88.

Wimmer, B., Craig, B.H., Pilkington, J.G., Pemberton, J.M., 2004. Non-invasive assessment of parasitic nematode species diversity in wild Soay sheep using molecular markers. Int J Parasitol 34, 625-631.

Zeng, Z.B., 1994. Precision mapping of quantitative trait loci. Genetics 136, 1457-1468.

Zhang, X.S., Hill, W.G., 2002. Joint effects of pleiotropic selection and stabilizing selection on the maintenance of quantitative genetic variation at mutation-selection balance. Genetics 162, 459-471.

Zhang, Y.M., Mao, Y., Xie, C., Smith, H., Luo, L., Xu, S., 2005. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (Zea mays L.). Genetics 169, 2267-2275.

Appendix

Here below are papers published or in preparation which contain results produced and elaborated during this PhD project (see also page 6 for list of publications).

DOI: 10.1534/genetics.106.057141

Development of a Linkage Map and Mapping of Phenotypic Polymorphisms in a Free-Living Population of Soay Sheep (*Ovis aries*)

Dario Beraldi,*,[1] Allan F. McRae,*,[†] Jacob Gratten,[‡] Jon Slate,[‡] Peter M. Visscher*,[†] and Josephine M. Pemberton*

*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom, [†]Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane 4029, Australia and [‡]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom

Manuscript received February 16, 2006
Accepted for publication April 7, 2006

ABSTRACT

An understanding of the determinants of trait variation and the selective forces acting on it in natural populations would give insights into the process of evolution. The combination of long-term studies of individuals living in the wild and better genomic resources for nonmodel organisms makes achieving this goal feasible. This article reports the development of a complete linkage map in a pedigree of free-living Soay sheep on St. Kilda and its application to mapping the loci responsible for three morphological polymorphisms for which the maintenance of variation demands explanation. The map was derived from 251 microsatellite and four allozyme markers and covers 3350 cM (~90% of the sheep genome) at ~15-cM intervals. Marker order was consistent with the published sheep map with the exception of one region on chromosome 1 and one on chromosome 12. *Coat color* maps to chromosome 2 where a strong candidate gene, tyrosinase-related protein 1 (TYRP1), has also been mapped. *Coat pattern* maps to chromosome 13, close to the candidate locus *Agouti*. *Horn type* maps to chromosome 10, a location similar to that previously identified in domestic sheep. These findings represent an advance in the dissection of the genetic diversity in the wild and provide the foundation for QTL analyses in the study population.

AN area of fundamental research in evolutionary genetics concerns the closely related issues of understanding the determinants of trait variation in natural populations and understanding how genetic variation for traits is maintained in the face of natural selection. The first of these problems is often summarized as the "genetic architecture" question: in general we would like to know whether genes of large effect commonly segregate in natural populations or whether the infinitesimal model, *i.e.*, that most traits are controlled by many genes of small effect, is appropriate—or, perhaps more likely, some configuration in between (Barton and Keightley 2002; Brem and Kruglyak 2005). Similarly, we would like to know to what extent genetic interactions such as dominance, pleiotropy, and epistasis contribute to the evolutionary dynamics of a population. The second problem was long ago identified by Fisher (1958): How is it that genetic variation for traits persists when selection is so often directional? The answer to this question must lie not only in the genetic architecture question, but also in the modes of selection and their temporal and spatial stability.

[1]*Corresponding author:* Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, UK. E-mail: dario.beraldi@ed.ac.uk

In principle, the arrival of abundant molecular markers, genetic maps, and whole-genome sequences allows us to address both genetic architecture and selection in much greater depth than ever before, since the role of variation at individual loci can be assessed. Mapping trait loci is a starting point for providing information on the genetic architecture of a trait in terms of the number of genes involved, relative effect, and mode of expression (Erickson *et al.* 2004; Slate 2005). In turn, this allows study of the relationship between phenotype and genotype and inference of the selective forces acting on the critical locus. Furthermore, by mapping genes, it is possible to test for the presence of gene-by-gene (epistatic) and gene-by-environment interactions, which are thought to contribute to phenotypic variation in natural and controlled settings (Carlborg and Haley 2004; Erickson 2005). In addition, the discovery of the map location of genes that influence phenotypic variation means that patterns of linkage disequilibrium (LD) and haplotype structure can be examined, which may provide insights about population history and selection. Unfortunately, some of the characteristics that make experimental populations so practical for linkage mapping also restrict the degree to which findings can be extrapolated to natural populations. Usually, geneticists generate segregating populations derived from one or a few pair of parents, which are often inbred and selected for the

extreme phenotypes. In addition, the population is raised in a uniform and controlled environment (*e.g.*, a greenhouse) where nongenetic sources of phenotypic variation are minimized. On the one hand, this strategy maximizes the power of analysis; *i.e.*, it increases the probability of finding a statistical association between marker genotype and phenotypic trait, but on the other hand, as the aim of genetic research becomes the understanding of how selection shapes genomes, the findings in experimental crosses are of limited applicability (Roff and Simons 1997; Conner 2002). In the wild, individuals are exposed to environmental and genetic forces (*e.g.*, genotype-by-environment interaction, pleiotropy, epistasis, maternal effects), some of which are unwittingly or deliberately diminished in experimental settings and may conceal important effects in the wild (Kroymann and Mitchell-Olds 2005; Wilson *et al.* 2005a,b). Although these forces are particularly difficult to detect in the wild, their possible absence from an experimental design may lead to biased conclusions.

A refined understanding of the process of evolution can be expected if the precise loci underlying trait variation can be identified and their behavior studied in free-living populations. Hence, a recent development is the application of genomic analyses to studies of free-living populations. Techniques for generating large numbers of genetic markers (*e.g.*, AFLPs and microsatellites) and the availability of markers from related model species means that genetic maps and quantitative trait locus (QTL) searches in organisms originally studied in the wild are becoming more common (Erickson *et al.* 2004; Slate 2005). To date, most of these studies have involved wild plants or animals brought into and bred in the laboratory. Although in some cases the experimental design makes use of pedigrees generated from several lines (Zhang *et al.* 2005) and investigates fitness-related traits (Laurie *et al.* 2004), such studies do not directly address the action of natural selection as the study organisms are the product of breeding programs. Other projects have been designed to answer specific ecological or evolutionary questions and to this end have employed individuals drawn from the wild and crossed under controlled conditions (Hawthorne and Via 2001; Lexer *et al.* 2003). The artificial development of the mapping populations, however, may generate genetic variation that may not occur in the wild (Erickson *et al.* 2004; Slate 2005). Given the existence of several studies of individually monitored, pedigreed individuals living in the wild, an obvious extension of these studies is to generate genetic maps and attempt to map genes underlying trait variation in nature. To date, however, we know of only two such studies pursuing this line (excluding studies of humans, where cultural factors make extension of findings to animal populations difficult). In red deer (*Cervus elaphus*) living on the island of Rum, Slate *et al.* (2002) obtained a partial map (~62% genome coverage) using microsatellite marker genotypes and then searched for QTL for a phenotypic trait, birth weight, finding three candidate regions for further investigation. Second, Hansson *et al.* (2005) have recently generated a preliminary genetic map (~25% genome coverage) for the great reed warbler (*Acrocephalus arundinaceus*) population at Lake Kvismaren, Sweden, again using microsatellites.

In this article, we describe the construction of a relatively much more complete genetic map for a free-living population, the Soay sheep (*Ovis aries*) living on St. Kilda, taking advantage of existing genomic resources available for domestic sheep. This population is the subject of a long-term, individual-based multidisciplinary study, which includes the collection of extensive phenotypic, ecological, and genetic information (Clutton-Brock and Pemberton 2004). Soay sheep are highly variable in appearance, with two independent polymorphisms of coat pigmentation (coat color and coat pattern) and polymorphic horns [normal, deformed ("scurred"), or polled horns]. Selection acting on two of these polymorphisms, coat color and horn type, has been previously documented (Milner *et al.* 2004). We demonstrate the utility of the genetic map by mapping the genes underlying these three polymorphic traits, setting the scene for better understanding of selection on these traits and for future QTL searches in the study population.

MATERIALS AND METHODS

Mapping population: The Soay sheep on the islands of Soay and Hirta (St. Kilda archipelago, northwestern Scotland, UK, 57°49′ N, 08°34′ W) are feral populations of a breed regarded as the most primitive in Europe (Campbell 1974; Doney *et al.* 1974); today, the sheep population of Hirta varies between 600 and 2000 individuals. Since 1985 regular expeditions have been sent to St. Kilda to monitor the population dynamics and to record the entire history of individuals living in Village Bay, Hirta (Clutton-Brock *et al.* 2004a). No predators are present on St. Kilda.

The mapping population analyzed in this study was selected from a larger Soay sheep data set comprising >3300 individuals with phenotypic records. In this population, maternity is determined by observation, and paternity is inferred through molecular analysis (Overall *et al.* 2005). The mating system is polygynous and promiscuous, such that very few full-sibs occur in the population. To trade off between power of linkage mapping and amount of genotyping work, we selected and genotyped half-sibships with 12 or more well-phenotyped individuals and their common parent, plus half-sibships with at least 10 animals that were related to previously selected animals. In addition, we included in the mapping pedigree file, but did not genotype, the noncommon parents and the ancestors of the half-sib families. Although not genotyped and in some cases phenotypically less well characterized, these additional individuals link different sibships, which otherwise would appear as unrelated. This strategy increases the number of informative meioses as missing genotypes and marker phases, in some cases, can be inferred from the knowledge that different individuals share the same ancestors. In total,

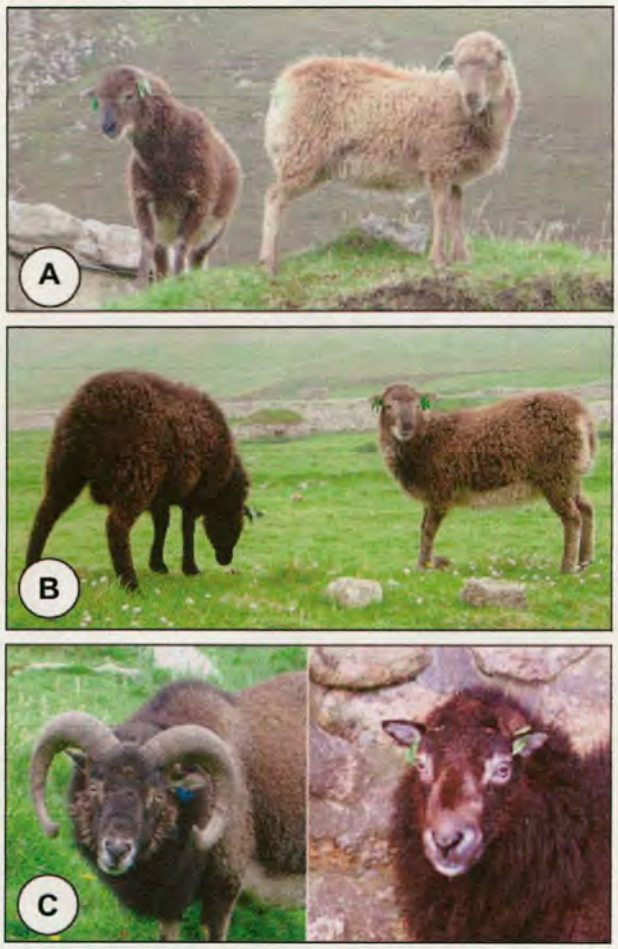

FIGURE 1.—Soay sheep showing the three traits subjected to linkage mapping. (A) Coat color polymorphism: dark (left) and light (right) lambs. (B) Coat pattern polymorphism: self (left) and wild-type (right) lambs; note lack of contrast in color between belly and rest of the body and the intensified coat color in the self individual. (C) Horn-type morphs in adult males: normal (left) and extreme scurred (right).

the mapping pedigree numbers 882 animals with 571 paternal links and 663 maternal links, of which 588 animals were genotyped (supplemental Figure S1 at http://www.genetics.org/supplemental/).

Polymorphic traits: In Soay sheep the color of the pelage is determined by two independently segregating polymorphisms, one affecting the color of the coat (hereafter referred to as coat color, locus *Coat color*, Figure 1A), and the other determining the contrast in color between belly and coat (hereafter referred to as coat pattern, locus *Coat pattern*, Figure 1B). Coat color can be classified into two distinct phenotypes, dark and light, which occur in a ratio of ~3:1. Segregation analyses in mainland Soays (DONEY *et al.* 1974) and in resolved pedigrees on St. Kilda (COLTMAN and PEMBERTON 2004) suggest that a single biallelic locus, in which dark is completely dominant to light, determines the two classes (see Table 1). With respect to coat pattern, Soay sheep with the "wild-type" morph have a paler belly and rump than the rest of the coat while the "self" morph is characterized by a uniform and more intense coat color. The wild and self morphs occur in a ratio of ~20:1. This variation is also determined by a single biallelic locus with wild type completely dominant to self (COLTMAN and PEMBERTON 2004). Wild-type sheep have hairs in which the dark color is alternated by pale bands, a pattern commonly found in wild mammals and usually due to the *Agouti* locus (BENNETT and LAMOREUX 2003). Conversely, in self sheep the hairs have no banding pattern (CLUTTON-BROCK *et al.* 2004b).

TABLE 1

Phenotypic distributions and underlying genotypes of the study traits in the genotyped members of the Soay mapping pedigree (maximum $n = 588$)

Trait (n)	Phenotype	Genotype	Frequency
Coat color (560)	Dark	*Dark/–*	0.74
	Light	*Light/Light*	0.26
Coat pattern (560)	Wild	*Wild/–*	0.94
	Self	*Self/Self*	0.06
Horn type—females (286)	Normal	Ho^+/Ho^+ Ho^+/Ho^L	0.38
	Scurred	Ho^L/Ho^L Ho^+/Ho^P	0.24
	Polled	Ho^P/Ho^P Ho^P/Ho^L	0.38
Horn type—males (270)	Normal	$Ho^+/–$ $Ho^L/–$	0.90
	Scurred	Ho^P/Ho^P	0.10

With respect to horn type (locus *Horn type*), Soay sheep are polymorphic for horns in both sexes. Females are classified into three horn types: normal (33%), scurred (vestigial and deformed, 28%), and polled (hornless, 39%), whereas in males only the normal (87%) and scurred (13%) phenotypes occur (Figure 1C, Table 1). Although the inheritance of the horn phenotype is not completely understood, pedigree data on St. Kilda are consistent with a single locus with three alleles (normal horned, sex-limited horned, and polled) showing sex-specific expression and dominance (COLTMAN and PEMBERTON 2004), a model originally proposed for Merino sheep (DOLLING 1961).

DNA extraction and microsatellite genotyping: Commercial kits were used to isolate DNA from blood samples (GFX genomic blood purification kit, Amersham Biosciences) or ear punches (GenomicPrep cells and tissue DNA isolation kit, Amersham Biosciences) following the manufacturer's instructions. When the amount of starting material was too small or degraded to allow the use of these methods, the DNA was extracted using Chelex resin beads (Chelex 100 Resin, Bio-Rad Laboratories, Hercules, CA). About 1–5 mg of blood or tissue was incubated at 56° overnight in 300 μl of a 5% Chelex and 0.1 μg/μl proteinase K solution followed by 5 min at 95°. Before PCR amplification, the DNA solution extracted with either method was diluted 1:4 with ddH_2O and 2 μl was air dried in 96-well PCR plates.

To construct a map with markers evenly distributed throughout the genome, the Australian Sheep Gene Mapping website (http://rubens.its.unimelb.edu.au/~jillm/jill.htm) was taken as a reference to select microsatellite markers on the basis of their location and information content. PCR amplifications were performed in 5 μl volume, and $MgCl_2$ concentration was adjusted between 1.5 and 4.0 mM to achieve optimal quality of the reaction. Two touchdown PCR programs were initially tested for each marker on a panel of eight sheep: one in which

the annealing temperature was progressively decreased during the first 10 of 29 cycles from 60° to 50°, and the other in which the decrease was from 65° to 55°. Fluorescent primers (6FAM, VIC, PET, or NED fluorescence) were synthesized by Applied Biosystems (Foster City, CA). Fragment lengths were analyzed on an ABI3730 DNA Analyzer and genotypes were determined using GeneMapper v3.0 (Applied Biosystems).

To estimate the genotyping error rate, 84–258 randomly chosen individuals were regenotyped at 10 loci with average polymorphism. Genotyping error rate was also assessed on the basis of mother–offspring mismatches using CERVUS 2.0 (MARSHALL *et al.* 1998).

Linkage map and genome scan: Parent–offspring genotype inconsistencies arising from incorrect paternity assignment (32 incorrect links found) or typing errors were detected through the program PedCheck (O'CONNELL and WEEKS 1998) and either resolved by rechecking the parentage records and genotypes or scored as untyped. Some cases of paternity mis-assignment were expected since in the original data set paternity was assigned with only 80% confidence (OVERALL *et al.* 2005).

Linkage mapping was performed using CRI-MAP v2.4 (GREEN *et al.* 1990) to determine the marker order, intermarker intervals, two-point LOD scores, and number of informative meioses. The complexity of the pedigree and the number of markers employed made a systematic testing of all the possible map combinations impractical. In most cases, the size of the pedigree did not allow the analysis of more than seven or eight markers at a time; therefore, sets of overlapping markers were tested sequentially until a whole chromosome was mapped. Markers expected to belong to the same chromosome were first input into CRI-MAP following the order reported in domestic sheep (Australian Sheep Gene Mapping website at http://rubens.its.unimelb.edu.au/~jillm/jill.htm). The log 10 likelihood of the initial marker order was then compared with that of alternative orders (*flips2* or *flips3* function) to detect more likely combinations (*i.e.*, higher log 10 likelihood). An increase in log 10 likelihood of three or more was considered evidence of a significantly more probable map (MORTON 1955). In cases of inconsistency between Soay and domestic sheep, the most probable Soay order was retained after having ruled out possible human or technical mistakes. Markers mapping to unexpected locations or supported by a weak LOD score (<1.8) were also tested for linkage (two-point function) against all the other markers in the database to detect whether better positions could be found.

Coat color and *Coat pattern* loci were mapped using CRI-MAP assuming each trait was encoded by a single locus with two alleles showing complete dominance: the *Dark* allele dominant over *Light* and the *Wild* dominant over *Self* (COLTMAN and PEMBERTON 2004; Table 1). A test for linkage between the target locus and any of the mapped markers was performed by means of the two-point function of CRI-MAP. The best position of a candidate locus was searched for by means of the *flips* and *fixed* functions of CRI-MAP to test alternative map positions. Consistent with the criteria used in the map construction, an increase in the log 10 likelihood of the map of three or more was taken as evidence of a significantly more likely position. In the case of *Coat pattern*, the low frequency of *Self* morphs resulted in a low number of informative meioses. To confirm or reject a suggestive linkage, more sheep in families segregating for *Coat pattern* were genotyped at markers in the relevant region (see RESULTS).

The *Horn type* locus was first investigated using CRI-MAP under a simplified model to scan the whole genome, and then the LINKAGE package (TERWILLIGER and OTT 1994) was employed to perform parametric multipoint analysis in target region(s) identified by the preliminary scan. In CRI-MAP, *Horn type* was coded as a single biallelic locus where the *Normal* (Ho^+) allele was dominant and the *Polled* (Ho^P) allele was recessive in males (Ho^+ allele conferring normal horns when heterozygote or homozygote and Ho^P allele resulting in scurred horns when homozygote), whereas in females Ho^+ and Ho^P alleles were expressed codominantly (normal, scurred, or polled horns given by Ho^+/Ho^+, Ho^+/Ho^P, and Ho^P/Ho^P, respectively). This model was simplified in that the presumptive *Sex-limited* allele (Ho^L) was not explicitly considered (see Table 1); modeling three alleles in CRI-MAP would have resulted in too many missing genotypes since this program does not allow a trait (or a marker) phenotype to be coded by more than one genotype. Although this simplification reduces the power of analysis, it does not bias the results.

Computational constraints due to the size of the pedigree and the number of inbreeding loops prevented the processing of the entire pedigree by parametric multipoint linkage analysis. To circumvent this problem, the mapping panel was split into 39 unlinked families. For the parametric multipoint analysis in LINKAGE (TERWILLIGER and OTT 1994), each sheep was assigned to one of five liability classes on the basis of their horn phenotype and sex: three classes for females (normal if Ho^+/Ho^+ or Ho^+/Ho^L, scurred if Ho^L/Ho^L or Ho^+/Ho^P, and polled if Ho^L/Ho^P or Ho^P/Ho^P; Table 1) and two classes for males (scurred horns if Ho^P/Ho^P, normal horns otherwise; males do not express the polled condition; Table 1). Finally, a sixth (fictitious) class was assigned to animals without phenotypic information; the underlying genotypes were assumed to have complete penetrance.

Horn allele frequencies were taken from COLTMAN AND PEMBERTON (2004) as 0.441, 0.170, and 0.389 for Ho^+, Ho^L, and Ho^P, respectively. Marker allele frequencies were estimated from the whole pedigree by 100,000 Markov chain Monte Carlo iterations implemented in Loki (HEATH 1997); this procedure is based on a stochastic process and as such does not provide an exact result, but allows the handling of very large and complex pedigrees.

The LOD threshold of 3.3 to declare evidence of linkage corresponds to the value usually applied to human pedigrees (LANDER and KRUGLYAK 1995). This decision was taken on the basis that the sizes of the sheep and human linkage maps are comparable.

RESULTS

Soay sheep linkage map: The Soay sheep linkage map was developed with 247 microsatellite and four allozyme markers, giving a total of ~124,000 genotypes, which generated a map with ~15-cM intermarker spacing across 3350 cM, equivalent to ~3080 cM on the International Mapping Flock (IMF) map and corresponding to ~90% of the sheep genome. Figure 2 compares the Soay sheep linkage map with the domestic sheep map (MADDOX *et al.* 2001); the APPENDIX lists the mapped markers and their characteristics. The mean number of alleles per locus was 4.6 with a mean polymorphism information content (PIC) of 0.52, which are lower values than those recorded on the Australian Sheep Gene Mapping website (http://rubens.its.unimelb.edu.au/~jillm/jill.htm) for the same markers typed in the IMF (10 alleles and PIC = 0.75); this is perhaps not surprising since the latter figures are for a pedigree derived from several sheep breeds. On average, each

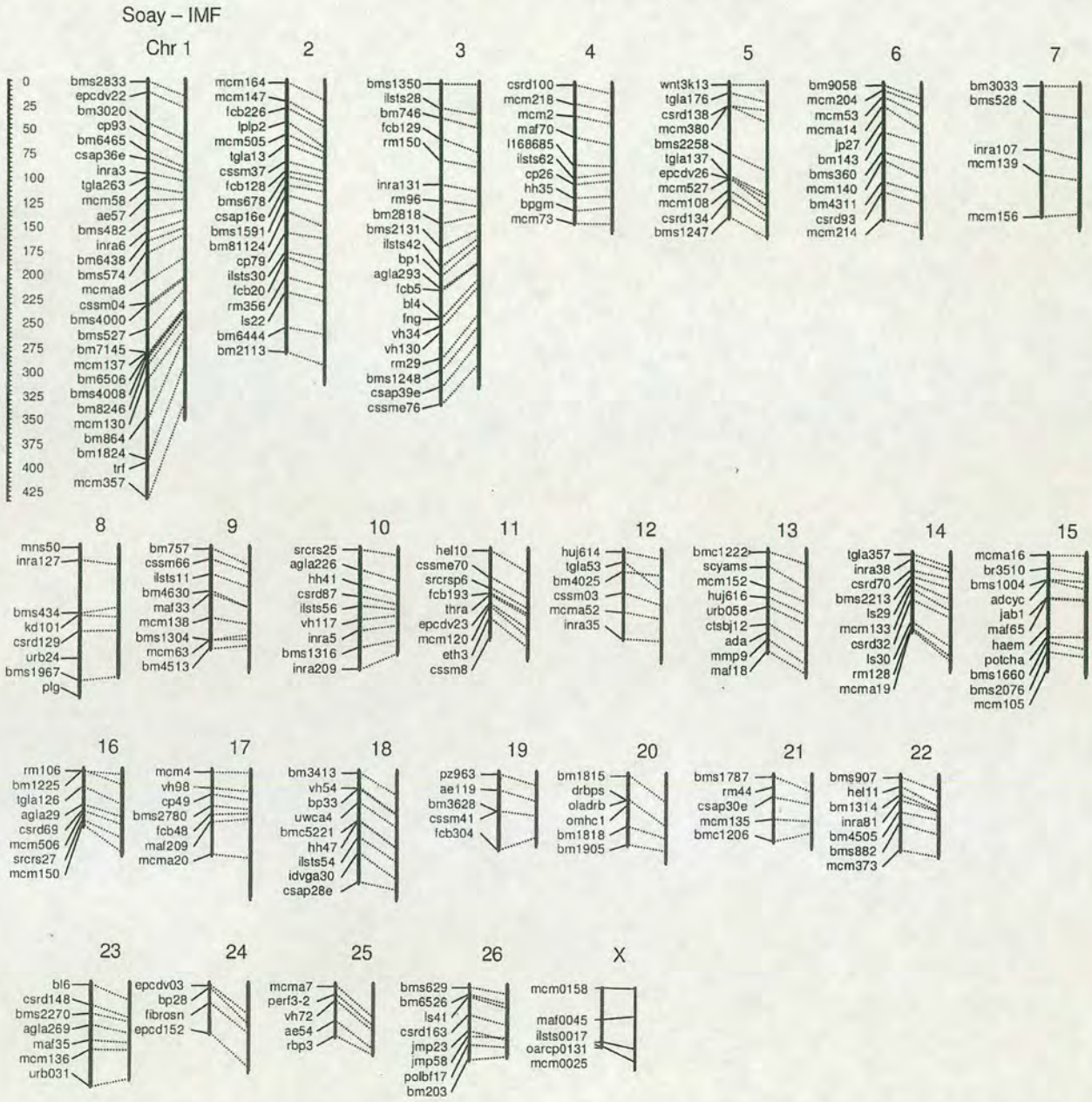

FIGURE 2.—Soay sheep map compared with the domestic (IMF) sheep map v4.3 (Australian Sheep Gene Mapping at http://rubens.its.unimelb.edu.au/~jillm/jill.htm). In each pair, the Soay chromosome is on the left side; dotted lines connect markers shared by both maps. The ruler at the top left corner represents a centimorgan scale.

marker was typed in 510 sheep (86% of the 588 sheep selected for genotyping) and generated 310 informative meioses. Genomewide, the mean LOD score for linkage between two adjacent markers was 14.6. Twenty-two marker intervals were linked with a LOD score of <2, but their marker positions were retained since they were in agreement with the domestic sheep map (MADDOX *et al.* 2001). Marker order was checked by means of the *flips* function of CRI-MAP and was consistent between the Soay and domestic sheep map in all but two cases: one on chromosome 1, where there is evidence for varying gene order in different sheep breeds (MCRAE and BERALDI 2006), and the other on chromosome 12, which we have not investigated further. Of the 1290 duplicated genotypes, 2.4% showed inconsistency with the first screening. The error rate based on mother–offspring mismatching was 1.49% as estimated by CERVUS (MARSHALL *et al.* 1998).

Linkage mapping of *Coat color*, *Coat pattern*, and *Horn type* loci: The phenotypic distributions of *Coat color*, *Coat*

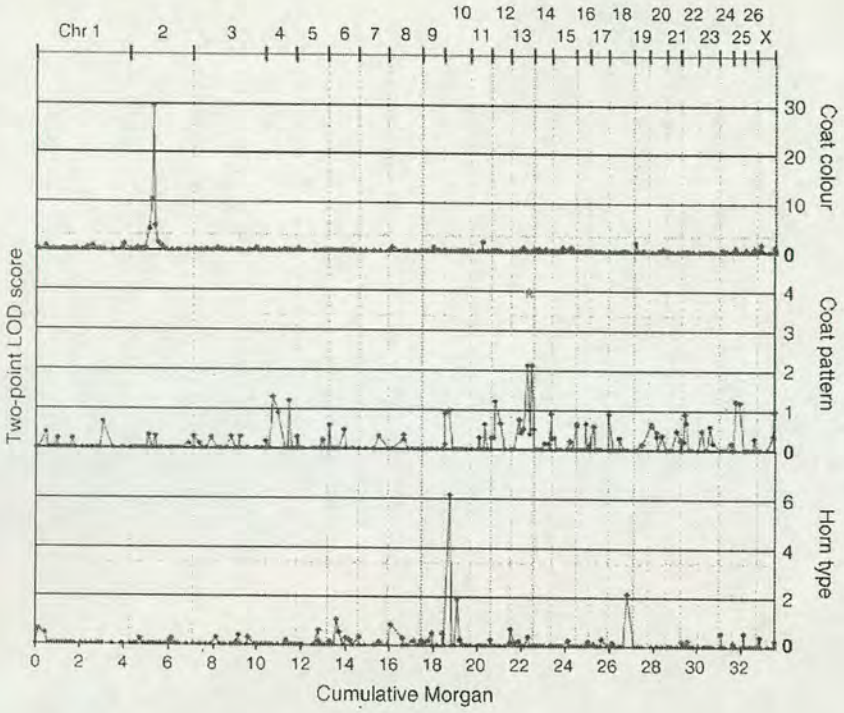

FIGURE 3.—Two-point LOD score profiles for linkage between study traits (on the right side of the graph) and markers (data points) in the Soay sheep map. The *x*-axis represents the cumulative map distance of the genome (morgan) with chromosome boundaries marked on top of the graph and by dotted vertical lines. *y*-Axes report the LOD scale. The dotted lines denote the theoretical genomewide significance threshold (LOD = 3.3). The asterisk in the middle panel shows the LOD score after having typed additional animals at marker CTSBJ12 (see text).

pattern, and *Horn type* in the mapping panel are reported in Table 1; these proportions do not differ significantly from the entire Soay sheep data set (χ^2 test $P > 0.1$).

Coat color: The highest LOD score for linkage was found with BMS678 (two-point LOD = 29.5 at 0 cM), a microsatellite located on chromosome 2 (Figure 3). Other markers on chromosome 2 were significantly linked to the target locus, namely FCB128 (LOD = 10.4), CSAP16E (LOD = 5.1), and CSSM37 (LOD = 4.4), whereas none of the other markers in the Soay sheep map produced a significant result for linkage (LOD <2). Figure 4A shows in detail the best position for the *Coat color* locus in the map of chromosome 2; any other map order results in a significant decrease (>3) in the log 10 likelihood of the map.

Coat pattern: The highest linkage score (LOD = 2.1) was detected on chromosome 13 (Figure 3). This LOD score fell short of genomewide significance, but this is likely to be a consequence of the low frequency of the self morph (6%), which meant that the *Coat pattern* locus was segregating in only a few families and there were few informative meioses for mapping ($N = 32$). To confirm or reject this suggestive linkage, another 78 animals composing 15 families segregating for coat pattern were genotyped for the two microsatellite markers encompassing the LOD score peak (CTSBJ12 and MMP9), and the association between marker CTSBJ12 and the *Coat pattern* locus rose to LOD = 3.9 with no recombinants between these loci (Figures 3 and 4B).

Horn type: Consistent with MONTGOMERY *et al.* (1996), CRI-MAP detected linkage between *Horn type* and AGLA226 on chromosome 10 (LOD = 6.1, Figure 3), but no other marker on chromosome 10 or elsewhere in the genome showed any significant linkage. Once the best location for *Horn type* was established on the chromosome 10 map (by use of the *fixed* function), CRI-MAP positioned *Horn type* distal to SRCRS25, the most telomeric marker on chromosome 10 (21.1 cM away from AGLA226). However, the likelihood of *Horn type* at this position was not significantly greater than in the interval between AGLA226 and SRCRS25 (log 10 likelihood: −150.16 *vs.* −151.44), although significantly better than in the interval AGLA226–HH41 (log 10 likelihood: −153.17). Therefore, at this stage we concluded that *Horn type* is located on chromosome 10 distal to or in the vicinity of AGLA226 (Figure 4C), but an accurate map position could not be assigned.

As described in MATERIALS AND METHODS, the CRI-MAP model of *Horn type* is simplified and does not account for the *Horn type* and marker allele frequencies. Therefore, the analysis was improved by performing multipoint parametric mapping to derive a more accurate estimate of the *Horn type* locus position. The *Horn type* locus was tested for linkage against AGLA226 and its two flanking markers. The LOD profile found by the multipoint analysis (Figure 5) suggests that the 1-LOD support interval for the presence of the target locus spans ~16 cM.

DISCUSSION

As a step toward the comprehension of the genetic dynamics of wild populations, this article reports the development of a genetic map in a free-living population,

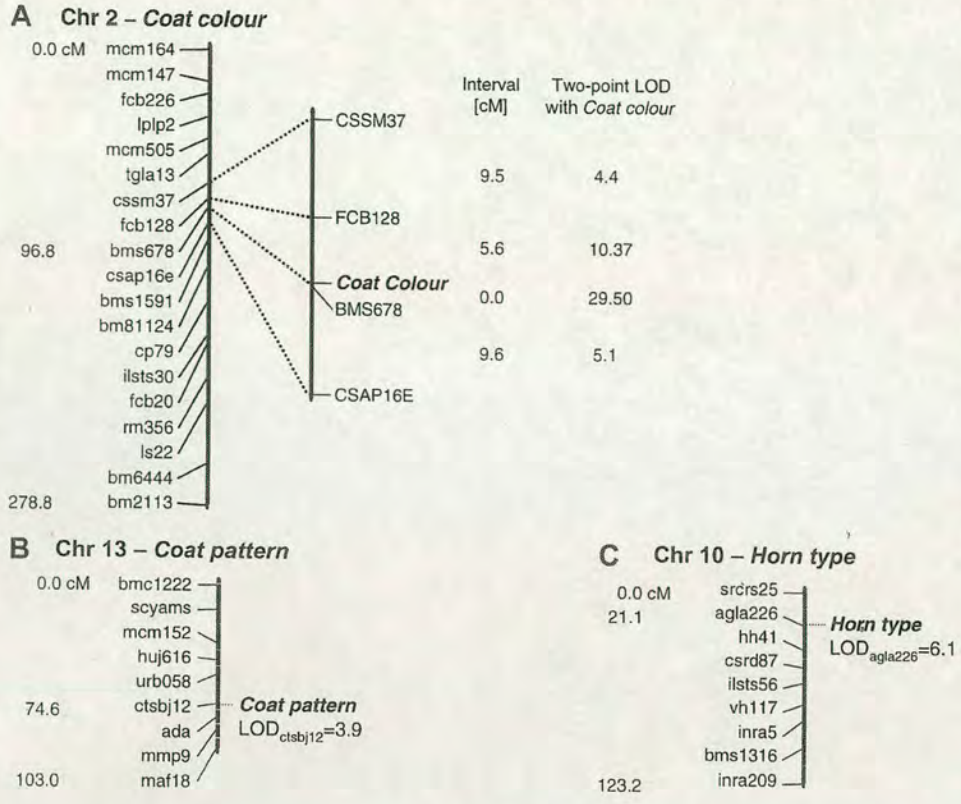

FIGURE 4.—Target regions identified by the genome scan for the three study traits. (A) Chromosome 2 full map and detailed map of the region carrying the *Coat color* locus. (B) Suggestive region for the *Coat pattern* locus on chromosome 13. (C) The *Horn type* location detected on chromosome 10 in the vicinity of AGLA226.

the Soay sheep on St. Kilda, and its use in a genome scan to map the loci responsible for three morphological traits. To the best of our knowledge, this is one of the first accomplishments of gene mapping in a free-living population. The Soay sheep on St. Kilda present interesting features from an evolutionary and genetic point of view: their number is naturally regulated by a combination of food availability, parasite burden, and winter weather (COULSON *et al.* 2001; CLUTTON-BROCK *et al.* 2004a; WILSON *et al.* 2004), factors that, together, cause substantial fluctuations in population size (COULSON *et al.* 2001; CLUTTON-BROCK *et al.* 2004a).

Development of the Soay sheep linkage map: The map presented here has been developed with the primary purpose of localizing genes of evolutionary interest. The map position of a locus can integrate extant models to describe the population dynamics of Soay sheep. Especially when the phenotype conveys little information about the underlying genotype, as is the case for many quantitative characters, the monitoring of the target trait is improved and complemented by the genotype inferred through linked markers.

Patterns of allelic association in terms of linkage disequilibrium and population structure provide insights into history and selection of a population (ABECASIS *et al.* 2005). To this end, a linkage map is a starting point to enrich regions of interest with markers to assess the extension of the association and to compare the latter with theoretical expectations (MCRAE *et al.* 2005). Genomic tools such as comparative mapping will facilitate the discovery of additional markers and candidate genes in target regions.

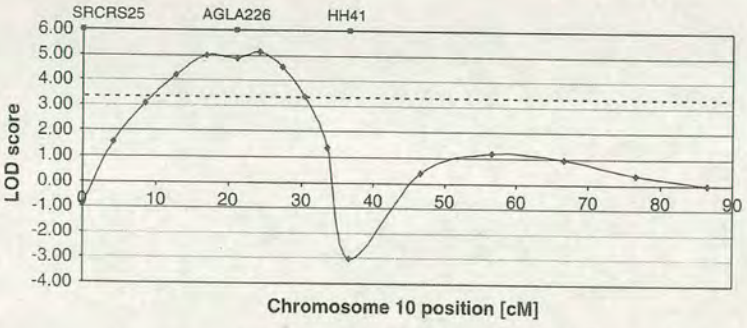

FIGURE 5.—Parametric four-point mapping of the *Horn type* locus. AGLA226 on chromosome 10, the marker showing the strongest two-point linkage in CRI-MAP, and two adjacent markers (SRCRS25 and HH41) were simultaneously tested against *Horn type*. The location of the three markers is shown at the top. The *Horn type* position was tested every 5 cM (data point). The dashed line denotes the theoretical genomewide significance threshold (LOD = 3.3).

Mapping of *Coat color*, *Coat pattern*, and *Horn type*: The attempt to map the locus responsible for coat color variation successfully yielded a region on chromosome 2 (Figure 3) defined by a window of ~15 cM (Figure 4A) in which the *Coat color* locus cosegregates with BMS678. Independently, J. Gratten, D. Beraldi, B. Lowder, A. McRae, P. Visscher, J. Pemberton and J. Slate (unpublished results), following a candidate gene method, have tested for association with different genes known to affect coat color in mammals and have identified the responsible gene (TYRP1) and its causal mutation.

The interest in coat color in Soay sheep stems from the differential survival between dark and light animals although no predators are present on St. Kilda and no obvious environmental conditions should favor one color over the other. It has previously been found that dark coats are positively selected during some high-mortality winters, but this is inconsistent and in other winters selection favors light-colored sheep or neither morph (Moorcroft *et al.* 1996; Milner *et al.* 2004). Dark animals are significantly heavier than light ones, providing a possible mechanism for their better survival (Clutton-Brock *et al.* 1997). There is no difference in female fecundity between dark and light sheep (Clutton-Brock *et al.* 1997). At present, there is no explanation for why the light-color morph is maintained in the population; clearly, being able to distinguish the three genotypes may shed light on this puzzle. Hypotheses and future work to explain the difference in survival will take advantage of the map position and molecular characterization of the *Coat color* gene. A comparison between LD in the FCB128–CSAP16E interval and background LD in the Soay sheep genome should also provide information about the origin and evolutionary consequences of coat color variation.

With respect to *Coat pattern*, the high frequency of the wild morph (94% of the sheep scored, Table 1) severely reduced the number of informative meioses (32) so that strong linkage to any marker was unlikely to be found. The power of linkage mapping is proportional to the fraction of parents heterozygous at both the target locus and the linked markers. This combination generates the necessary marker-trait association in the progeny (Lynch and Walsh 1998). It follows that if the target locus has a highly skewed allelic distribution, few heterozygous individuals are generated and more meioses need to be scored (the information content, estimated as PIC, reaches the highest value when all the alleles have the same frequency). Accordingly, the highest LOD score for *Coat pattern* reached only 2.1 on chromosome 13 (Figure 3) after an initial scan. However, the extension of the sample size confirmed this suggestive linkage. Interestingly, chromosome 13 harbors the *Agouti* locus, a candidate for *Coat pattern* (Parsons *et al.* 1999). *Agouti* encodes for an antagonist of the melanocortin receptor, causing a switch from eumelanin to pheomelanin production in the pigment-producing cells, which results in the characteristic banding pattern observed in Soay sheep hairs and other mammals (Bennett and Lamoreux 2003). To date, we have not detected selection acting on the *Coat pattern* locus.

Multipoint parametric linkage analysis was not performed for *Coat color* and *Coat pattern* because, in contrast to *Horn type*, the CRI-MAP model for *Coat color* and *Coat pattern* was already consistent with the most likely model, so that little or no improvement would have been gained by multipoint parametric analysis.

The mapping of *Horn type* returned a telomeric region on chromosome 10 previously detected by Montgomery *et al.* (1996; Figures 4C and 5). This work opens the way for multiple strategies to fine map and isolate the *Horn type* gene. These include exploitation of bioinformatic tools to enrich the target region with SNPs and other microsatellites and identification of positional candidates by comparison with the annotated genome assemblies of cattle and other species. Like coat color, horn phenotype is under selection in Soay sheep and other wild populations. In ruminants, horns are typically used in intrasexual conflict, particularly among males where they reach much greater size. Previous analyses of Soay sheep have suggested that normal-horned males and scurred females have the highest annual breeding success (Clutton-Brock *et al.* 1997; Stevenson *et al.* 2004), but that in winters characterized by high mortality, the scurred phenotype is generally favored in both sexes (Moorcroft *et al.* 1996). Exactly how these forces maintain variation in the population is the subject of current research and would clearly be helped by being able to distinguish individuals by genotype rather than by phenotype. Therefore, the *Horn type* region is an attractive target for molecular evolution studies.

Future directions: The traits analyzed here are characterized by relatively simple inheritance patterns which, to some extent, may limit their applicability to the understanding of the process of evolution. However, this project opens the way to the more challenging task of detecting QTL affecting a variety of morphological and physiological traits. The Soay sheep has been the subject of a number of studies aimed at estimating quantitative genetic parameters for traits like birth weight and body size (Coltman *et al.* 1999; Milner *et al.* 2000, 2004). It has been found that the additive genetic variance of these traits is low but not null, despite the pressure of selection acting on them (Milner *et al.* 2000). As these previous studies have been conducted under the infinitesimal model framework, the dissection of these traits through QTL mapping to determine eventual Mendelian factors would represent a major breakthrough toward the comprehension of the evolutionary processes in the wild.

We thank J. G. Pilkington and many volunteers for collecting field data and genetic samples; the National Trust for Scotland for granting permission to work on St. Kilda and QinetiQ for logistical support; and two anonymous reviewers for their comments on the first version of

this article. The long-term data collection on St. Kilda has been supported by Natural Environment Research Council (NERC) and Wellcome Trust grants to T. H. Clutton-Brock, B. T. Grenfell, L. E. B. Kruuk, M. J. Crawley, and J.M.P. This study was funded by NERC through its Environmental Genomics thematic program (grant no. NER/T/S/2002/00189).

LITERATURE CITED

Abecasis, G. R., D. Ghosh and T. E. Nichols, 2005 Linkage disequilibrium: ancient history drives the new genetics. Hum. Hered. **59:** 118–124.

Barton, N. H., and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nat. Rev. Genet. **3:** 11–21.

Bennett, D. C., and M. L. Lamoreux, 2003 The color loci of mice—a genetic century. Pigment Cell Res. **16:** 333–344.

Brem, R. B., and L. Kruglyak, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. USA **102:** 1572–1577.

Campbell, R. N., 1974 *St. Kilda and Its Sheep. Island Survivors: The Ecology of the Soay Sheep of St. Kilda.* The Athlone Press, London.

Carlborg, O., and C. S. Haley, 2004 Epistasis: Too often neglected in complex trait studies? Nat. Rev. Genet. **5:** 618–625.

Clutton-Brock, T. H., and J. M. Pemberton, 2004 *Soay Sheep Dynamics and Selection in an Island Population.* Cambridge University Press, Cambridge, UK.

Clutton-Brock, T. H., K. Wilson and I. R. Stevenson, 1997 Density-dependent selection on horn phenotype in Soay sheep. Philos. Trans. R. Soc. Lond. B Biol. Sci. **352:** 839–850.

Clutton-Brock, T. H., B. T. Grenfell, T. Coulson, A. D. C. MacColl, A. W. Illius *et al.*, 2004a Population dynamics in Soay sheep, pp. 52–88 in *Soay Sheep Dynamics and Selection in an Island Population*, edited by T. H. Clutton-Brock and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

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

Coltman, D. W., and J. M. Pemberton, 2004 Inheritance of coat colour and horn type in Soay sheep, pp. 321–327 in *Soay Sheep Dynamics and Selection in an Island Population*, edited by T. H. Clutton-Brock and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Coltman, D. W., J. A. Smith, D. R. Bancroft, J. Pilkington, A. D. MacColl *et al.*, 1999 Density-dependent variation in lifetime breeding success and natural and sexual selection in Soay rams. Am. Nat. **154:** 730–746.

Conner, J. K., 2002 Genetic mechanisms of floral trait correlations in a natural population. Nature **420:** 407–410.

Coulson, T., E. A. Catchpole, S. D. Albon, B. J. Morgan, J. M. Pemberton *et al.*, 2001 Age, sex, density, winter weather, and population crashes in Soay sheep. Science **292:** 1528–1531.

Dolling, C. H. S., 1961 Hornedness and polledness in sheep. IV. Triple alleles affecting horn growth in the Merino. Aust. J. Agric. Res. **12:** 353–361.

Doney, J. M., M. L. Ryder, R. G. Gunn and P. Grubb, 1974 *Colour, Conformation, Affinities, Fleece and Patterns of Inheritance of the Soay Sheep. Island Survivors: The Ecology of the Soay Sheep of St. Kilda.* The Athlone Press, London.

Erickson, D., 2005 Quantitative trait loci: mapping the future of QTL's. Heredity **95:** 417–418.

Erickson, D. L., C. B. Fenster, H. K. Stenoien and D. Price, 2004 Quantitative trait locus analyses and the study of evolutionary process. Mol. Ecol. **13:** 2505–2522.

Fisher, R. A., 1958 *The Genetical Theory of Natural Selection.* Dover, New York.

Green, P., K. Falls and S. Crooks, 1990 Documentation for CRI-MAP. Washington University, St Louis.

Hansson, B., M. Akesson, J. Slate and J. M. Pemberton, 2005 Linkage mapping reveals sex-dimorphic map distances in a passerine bird. Proc. Biol. Sci. **272:** 2289–2298.

Hawthorne, D. J., and S. Via, 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature **412:** 904–907.

Heath, S. C., 1997 Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am. J. Hum. Genet. **61:** 748–760.

Kroymann, J., and T. Mitchell-Olds, 2005 Epistasis and balanced polymorphism influencing complex trait variation. Nature **435:** 95–98.

Lander, E., and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. **11:** 241–247.

Laurie, C. C., S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush *et al.*, 2004 The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics **168:** 2141–2155.

Lexer, C., M. E. Welch, J. L. Durphy and L. H. Rieseberg, 2003 Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of Helianthus paradoxus, a diploid hybrid species. Mol. Ecol. **12:** 1225–1235.

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

Maddox, J. F., K. P. Davies, A. M. Crawford, D. J. Hulme, D. Vaiman *et al.*, 2001 An enhanced linkage map of the sheep genome comprising more than 1000 loci. Genome Res. **11:** 1275–1289.

Marshall, T. C., J. Slate, L. E. Kruuk and J. M. Pemberton, 1998 Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. **7:** 639–655.

McRae, A. F., and D. Beraldi, 2006 Examination of a region showing linkage map discrepancies across sheep breeds. Mamm. Genome **17:** 346–353.

McRae, A. F., J. M. Pemberton and P. M. Visscher, 2005 Modeling linkage disequilibrium in natural populations: the example of the Soay sheep population of St. Kilda, Scotland. Genetics **171:** 251–258.

Milner, J. M., J. M. Pemberton, S. Brotherstone and S. D. Albon, 2000 Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach. J. Evol. Biol. **13:** 804–813.

Milner, J. M., S. D. Albon, L. E. B. Kruuk and J. M. Pemberton, 2004 Selection on phenotype, pp. 190–216 in *Soay Sheep Dynamics and Selection in an Island Population*, edited by T. H. Clutton-Brock and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Montgomery, G. W., H. M. Henry, K. G. Dodds, A. E. Beattie, T. Wuliji *et al.*, 1996 Mapping the Horns (Ho) locus in sheep: a further locus controlling horn development in domestic animals. J. Hered. **87:** 358–363.

Moorcroft, P. R., S. D. Albon, J. M. Pemberton, I. R. Stevenson and T. H. Clutton-Brock, 1996 Density-dependent selection in a fluctuating ungulate population. Proc. R. Soc. Lond. B Biol. Sci. **263:** 31–38.

Morton, N. E., 1955 Sequential tests for the detection of linkage. Am. J. Hum. Genet. **7:** 277–318.

O'Connell, J. R., and D. E. Weeks, 1998 PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am. J. Hum. Genet. **63:** 259–266.

Overall, A. D., K. A. Byrne, J. G. Pilkington and J. M. Pemberton, 2005 Heterozygosity, inbreeding and neonatal traits in Soay sheep on St. Kilda. Mol. Ecol. **14:** 3383–3393.

Parsons, Y. M., M. R. Fleet and D. W. Cooper, 1999 The Agouti gene: a positional candidate for recessive self-colour pigmentation in the Australian merino. Aust. J. Agric. Res. **50:** 1099–1103.

Roff, D. A., and A. M. Simons, 1997 The quantitative genetics of wing dimorphism under laboratory and "field" conditions in the cricket Gryllus pennsylvanicus. Heredity **78:** 235–240.

Slate, J., 2005 Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Mol. Ecol. **14:** 363–379.

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

Stevenson, I. R., P. Marrow, B. T. Preston, J. M. Pemberton and K. Wilson, 2004 Adaptive reproduction strategies, pp. 243–275 in *Soay Sheep Dynamics and Selection in an Island Population*, edited

by T. H. Clutton-Brock and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Terwilliger, J. D., and J. Ott, 1994 *Handbook of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore.

Wilson, A. J., D. W. Coltman, J. M. Pemberton, A. D. Overall, K. A. Byrne *et al.*, 2005a Maternal genetic effects set the potential for evolution in a free-living vertebrate population. J. Evol. Biol. **18:** 405–414.

Wilson, A. J., J. G. Pilkington, J. M. Pemberton, D. W. Coltman, A. D. Overall *et al.*, 2005b Selection on mothers and offspring: Whose phenotype is it and does it matter? Evolution Int. J. Org. Evolution **59:** 451–463.

Wilson, K., B. T. Grenfell, J. G. Pilkington, H. E. G. Boyd and F. M. D. Gulland, 2004 Parasites and their impact, pp. 52–88 in *Soay Sheep Dynamics and Selection in an Island Population*, edited by T. H. Clutton-Brock and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Zhang, Y. M., Y. Mao, C. Xie, H. Smith, L. Luo *et al.*, 2005 Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics **169:** 2267–2275.

Communicating editor: M. Nordborg

APPENDIX

Details of the markers included in the Soay sheep genetic map

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	*N*[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l]		
												Pos[d]	No. all[g]	PIC[j]
1	BMS2833	40.43	11	0	533	496	5	0.762	0.743	0.7	0.0053	14.2	9	0.7
1	EPCDV22	10.17	32	11	543	463	6	0.773	0.743	0.707	0	26.4	7	NA
1	BM3020	24.99	13.6	43	481	440	6	0.778	0.785	0.75	0.0052	59.5	13	0.74
1	CP93	16.03	15.7	56.6	545	343	4	0.618	0.593	0.55	0	73.6	6	0.58
1	BM6465	42.6	8.2	72.3	522	405	6	0.707	0.737	0.697	0.0107	87.7	8	NA
1	CSAP36E	10.8	13.8	80.5	525	484	5	0.73	0.739	0.692	0.0058	93.5	10	0.82
1	INRA3	20.68	4.3	94.3	517	216	2	0.453	0.455	0.352	0	102.8	5	0.46
1	TGLA263	30.03	10	98.6	402	377	7	0.816	0.787	0.755	0	NA	NA	NA
1	MCM58	20.1	13.8	108.6	550	464	7	0.78	0.746	0.702	0	114	14	0.88
1	AE57	3.48	18.9	122.4	502	290	4	0.552	0.534	0.474	0.0545	120.7	9	0.76
1	BMS482	12.81	13.6	141.3	528	281	4	0.424	0.457	0.429	0.0707	132.3	11	NA
1	INRA6	12.53	8.7	154.9	500	267	3	0.582	0.563	0.466	0	142.7	8	0.52
1	BM6438	9.71	12.9	163.6	526	302	4	0.551	0.535	0.454	0	150.6	6	0.77
1	BMS574	4	28.6	176.5	434	238	4	0.562	0.581	0.518	0.0164	157.5	12	0.75
1	MCMA8	5.03	23.5	205.1	544	215	3	0.406	0.435	0.394	0	181.3	11	0.65
1	CSSM04	17.51	2.9	228.6	528	236	3	0.492	0.452	0.359	0	200.9	8	NA
1	BMS4000	3.47	23.6	231.5	425	202	5	0.595	0.569	0.496	0	204.1	15	0.82
1	BMS527	7.6	24.2	255.1	526	292	5	0.555	0.535	0.491	0.0117	214.9	12	0.82
1	BM7145	89.4	0.8	279.3	503	394	4	0.718	0.721	0.669	0	234.6	6	0.66
1	MCM137	69.12	2.1	280.1	558	523	7	0.763	0.752	0.715	0	233.4	15	0.85
1	BM6506	58.31	1.4	282.2	506	333	5	0.686	0.676	0.63	0.0075	235.6	6	NA
1	BMS4008	35.13	6.8	283.6	522	411	7	0.787	0.765	0.728	0	236.7	10	0.75
1	BM8246	10.2	15.6	290.4	525	350	5	0.617	0.599	0.558	0.009	242.2	9	NA
1	MCM130	0	40.7	306	553	363	4	0.571	0.541	0.435	0.0114	256.5	16	0.73
1	BM864	0	44.3	346.7	520	143	3	0.287	0.284	0.256	0	264.5	11	0.73
1	BM1824	40.75	3.1	391	501	322	5	0.653	0.671	0.621	0	294.1	5	0.68
1	TRF	0	36.8	394.1	539	437	7	0.785	0.764	0.728	0.0135	NA	NA	NA
1	MCM357	0		430.9	517	176	2	0.422	0.417	0.33	0.0631	332.6	11	0.85
2	MCM164	3.74	19.2	0	546	97	3	0.192	0.181	0.169	0	17.6	13	0.78
2	MCM147	8.58	6.8	19.2	533	414	7	0.717	0.722	0.68	0.0114	39.8	12	0.83
2	FCB226	3.49	14.7	26	529	127	4	0.386	0.377	0.348	0	44.6	12	0.79
2	LPLP2	4.35	13	40.7	546	441	6	0.749	0.739	0.696	0.0051	65.5	14	0.84
2	MCM505	3.15	10.2	53.7	551	80	5	0.181	0.182	0.176	0	71.4	8	0.7
2	TGLA13	8.47	17.8	63.9	397	277	4	0.718	0.737	0.687	0.0165	77.8	6	NA
2	CSSM37	30.66	9.5	81.7	545	374	5	0.556	0.566	0.537	0	91.1	12	0.54
2	FCB128	48.92	5.6	91.2	544	439	4	0.662	0.645	0.597	0.0241	99.4	8	0.72
2	BMS678	16.38	9.6	96.8	549	426	5	0.619	0.614	0.542	0	106.1	14	0.83
2	CSAP16E	6.24	9.8	106.4	523	146	2	0.283	0.285	0.244	0.0468	112.4	6	0.43
2	BMS1591	10.86	17.1	116.2	484	264	5	0.556	0.521	0.481	0.0131	127	16	0.85
2	BM81124	9.89	21.8	133.3	526	410	4	0.679	0.638	0.566	0	148.2	11	NA

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	*N*[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l]		
												Pos[d]	No. all[g]	PIC[j]
2	CP79	4.61	19.8	155.1	473	306	6	0.668	0.667	0.611	0.0248	161	8	NA
2	ILSTS30	8.73	5.4	174.9	550	120	2	0.213	0.21	0.188	0	182.4	10	0.7
2	FCB20	7.47	20.6	180.3	400	236	7	0.61	0.635	0.6	0.075	194	12	0.8
2	RM356	13.78	15.6	200.9	496	360	6	0.752	0.748	0.705	0.0244	211.1	16	NA
2	LS22	3.41	36.2	216.5	546	349	5	0.668	0.638	0.57	0	225.8	8	0.78
2	BM6444	7.72	26.1	252.7	521	425	6	0.758	0.755	0.713	0	260.2	13	0.88
2	BM2113			278.8	495	356	4	0.614	0.595	0.514	0.0218	291.9	8	0.67
3	BMS1350	4.76	25	0	438	187	5	0.393	0.403	0.377	0	0	12	NA
3	ILSTS28	15.04	9.9	25	553	356	5	0.633	0.607	0.56	0	32	14	0.82
3	BM746	5.17	21.9	34.9	537	287	3	0.529	0.537	0.43	0.0243	45.4	9	0.52
3	FCB129	3.49	22.7	56.8	528	347	4	0.595	0.625	0.577	0.0084	70.2	10	0.71
3	RM150	2.3	24.8	79.5	536	160	5	0.326	0.34	0.316	0	85.9	6	NA
3	INRA131	12.56	16.5	104.3	523	327	3	0.639	0.657	0.582	0.0278	111.4	7	NA
3	RM96	3.81	23.4	120.8	521	344	4	0.601	0.55	0.494	0.0129	126.7	6	0.56
3	BM2818	2.89	24.7	144.2	496	214	3	0.528	0.531	0.45	0.0726	134.9	5	NA
3	BMS2131	6.94	19.8	168.9	543	327	3	0.652	0.634	0.559	0.0753	150.4	5	NA
3	ILSTS42	15.04	9.9	188.7	530	365	7	0.675	0.654	0.612	0.007	159.1	10	NA
3	BP1	4.84	12.6	198.6	521	257	2	0.53	0.5	0.375	0	166	4	0.38
3	AGLA293	10.34	2.7	211.2	557	241	2	0.434	0.433	0.339	0.0525	184.7	4	0.36
3	FCB5	0.61	28.6	213.9	547	163	2	0.333	0.313	0.264	0	186.3	3	0.42
3	BL4	36.8	1	242.5	569	301	3	0.559	0.581	0.488	0.0095	202.5	8	NA
3	FNG	12.43	7.4	243.5	563	238	2	0.448	0.465	0.357	0	NA	NA	NA
3	VH34	0	30.3	250.9	572	329	5	0.579	0.576	0.532	0	210.7	9	0.68
3	RM29	0.75	11.7	281.2	177	69	5	0.644	0.592	0.53	0.0347	238.6	8	0.67
3	BMS1248	5.83	17.9	292.9	530	204	2	0.413	0.421	0.332	0.0588	251.1	11	0.67
3	CSAP39E	2.41	17.4	310.8	517	236	3	0.596	0.545	0.482	0.0761	268.7	11	0.75
3	CSSME76			328.2	542	106	3	0.231	0.236	0.22	0	290.4	9	0.68
4	CSRD100	9.49	19.5	0	549	207	3	0.364	0.364	0.309	0	9.2	8	0.71
4	MCM218	38.21	12	19.5	529	437	5	0.773	0.748	0.705	0.0052	26.5	9	0.82
4	MCM2	10.53	22.5	31.5	555	502	7	0.791	0.796	0.771	0.0038	39.4	11	0.8
4	MAF70	1.7	28.6	54	381	263	6	0.743	0.785	0.749	0.1455	61.4	17	0.9
4	L168685	7.76	12.9	82.6	462	220	4	0.545	0.559	0.503	0.0268	83.3	9	NA
4	ILSTS62	28.98	7	95.5	517	332	5	0.617	0.602	0.543	0	91.3	16	0.88
4	CP26	18.59	14.4	102.5	560	422	5	0.738	0.723	0.674	0.0054	98.7	6	0.74
4	HH35	21.57	13.3	116.9	548	408	4	0.695	0.696	0.643	0.0125	114.9	7	0.7
4	BPGM	15.12	12.5	130.2	545	394	6	0.653	0.641	0.611	0	126.1	7	0.7
4	MCM73			142.7	545	316	3	0.486	0.489	0.402	0.0851	143.6	13	0.88
5	WNT3K13	14.56	9.3	0	550	191	4	0.438	0.415	0.386	0	0	8	0.76
5	TGLA176	9.37	12.3	9.3	549	292	3	0.574	0.538	0.464	0	17.8	8	NA
5	MCM380	9.63	0	21.6	559	70	4	0.114	0.116	0.111	0	26.6	10	0.78

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	N[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l]		
												Pos[d]	No. all[g]	PIC[j]
5	CSRD138	0	50	21.6	535	383	5	0.6	0.577	0.519	0	38.7	10	0.76
5	BMS2258	5	22.7	71.6	551	266	2	0.454	0.479	0.364	0	89.6	8	NA
5	TGLA137	26.63	2.2	94.3	516	288	3	0.523	0.517	0.462	0	113.2	8	NA
5	EPCDV26	32.74	1.9	96.5	553	302	5	0.582	0.58	0.542	0	117.4	7	NA
5	MCM527	27.7	12.1	98.4	397	324	6	0.783	0.758	0.727	0.0152	125.5	6	0.67
5	MCM108	30.28	8.8	110.5	545	503	7	0.811	0.784	0.753	0	135.2	11	0.82
5	CSRD134	5.36	18.8	119.3	501	281	4	0.559	0.613	0.557	0.0286	140.8	6	0.55
5	BMS1247			138.1	545	370	3	0.662	0.629	0.552	0.0086	157	7	NA
6	BM9058	22.59	5	0	452	312	5	0.615	0.624	0.578	0	12.9	10	NA
6	MCM204	20.7	6.8	5	541	318	4	0.54	0.516	0.465	0	18.2	8	0.79
6	MCM53	6.29	10.7	11.8	558	378	5	0.67	0.664	0.623	0.0064	29.7	9	0.76
6	MCMA14	2.09	14.5	22.5	543	88	3	0.179	0.178	0.17	0	45	8	0.67
6	JP27	25.46	11.9	37	526	449	8	0.793	0.798	0.768	0.0084	NA	NA	NA
6	BM143	10.29	20.7	48.9	544	332	6	0.563	0.549	0.506	0	59	10	NA
6	BMS360	20.64	8.3	69.6	536	388	4	0.688	0.671	0.609	0	80.8	12	NA
6	MCM140	5.28	22.3	77.9	447	317	5	0.609	0.599	0.533	0.0123	95.8	10	0.79
6	BM4311	10.78	10.7	100.2	447	355	5	0.996	0.705	0.647	0	111.6	9	0.79
6	CSRD93	5.77	27.3	110.9	537	427	4	0.695	0.7	0.646	0	122.7	11	0.73
6	MCM214			138.2	552	417	4	0.679	0.658	0.597	0	147.1	7	0.72
7	BM3033	5.02	28.2	0	541	379	3	0.632	0.634	0.562	0.0084	0	17	NA
7	BMS528	1.7	38.2	28.2	533	348	3	0.612	0.618	0.536	0.0097	32.6	6	NA
7	INRA107	3.82	26.6	66.4	517	386	3	0.598	0.574	0.49	0.0358	75.4	17	0.79
7	MCM139	1.2	42.6	93	518	236	5	0.542	0.555	0.524	0.0203	97.6	11	0.81
7	MCM156			135.6	534	294	4	0.614	0.632	0.572	0.0756	133.4	8	0.77
8	MNS50A	10.11	13.3	0	547	463	5	0.779	0.737	0.688	0.0165	5	11	0.85
8	INRA127	0	54.6	13.3	503	211	3	0.342	0.366	0.301	0	17.2	8	0.81
8	BMS434	62.99	2.4	67.9	515	394	6	0.724	0.692	0.652	0.0069	61.3	9	NA
8	KD101	16.03	16.8	70.3	552	452	6	0.766	0.758	0.716	0.0048	71.1	12	0.83
8	CSRD129	2.99	27.2	87.1	541	328	5	0.636	0.627	0.566	0	86	11	0.82
8	URB024	9.86	23.5	114.3	528	270	4	0.532	0.549	0.49	0	117.5	12	0.65
8	BMS1967	6.09	16.1	137.8	535	503	6	0.776	0.769	0.731	0	132.8	11	0.71
8	PLG			153.9	277	144	2	0.596	0.5	0.375	0	NA	NA	NA
9	BM757	26.84	10.9	0	550	364	4	0.675	0.67	0.615	0.0069	16.2	7	NA
9	CSSM66	12.42	15.4	10.9	550	325	4	0.578	0.569	0.507	0.0309	24.2	10	NA
9	ILSTS11	9.6	18.2	26.3	533	290	3	0.465	0.449	0.401	0	40.1	10	0.68
9	BM4630	6.18	2.6	44.5	540	368	6	0.548	0.611	0.555	0.0451	60	8	0.78
9	MAF33	0	24.4	47.1	193	36	4	0.285	0.302	0.286	0.2015	60	9	0.7
9	MCM138	1.12	23.1	71.5	539	222	2	0.341	0.418	0.33	0.1234	78.5	9	0.64
9	BMS1304	3.31	0.4	94.6	45	33	6	1	0.782	0.737	0	89.6	12	NA
9	MCM63	29.71	8	95	546	362	5	0.632	0.645	0.602	0.016	93.1	14	0.87

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	N[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l] Pos[d]	IMF map[l] No. all[g]	IMF map[l] PIC[j]
9	BM4513			103	429	403	7	0.811	0.806	0.776	0	100.3	10	0.82
10	SRCRS25	7.22	21.1	0	553	253	4	0.483	0.498	0.463	0.0515	5.3	11	0.82
10	AGLA226	10.64	15.5	21.1	481	312	6	0.636	0.68	0.646	0.0078	31.5	8	0.78
10	HH41	9.83	11.7	36.6	545	366	7	0.615	0.634	0.585	0	45.9	11	0.8
10	CSRD87	14.02	9.8	48.3	538	291	4	0.61	0.585	0.504	0	56.2	10	0.78
10	ILSTS56	22.31	14.7	58.1	500	359	6	0.682	0.72	0.687	0.0621	61.2	9	0.7
10	VH117	19.16	10.7	72.8	541	451	6	0.726	0.734	0.693	0.0108	68.2	11	0.75
10	INRA5	11.47	17.8	83.5	403	314	8	0.695	0.679	0.635	0.0181	78	13	0.87
10	BMS1316	5.93	21.9	101.3	547	452	5	0.682	0.673	0.609	0	93.5	13	0.81
10	INRA209			123.2	546	179	2	0.264	0.297	0.253	0.043	108	10	0.58
11	HEL10	4.2	27.7	0	550	349	5	0.645	0.624	0.588	0.0077	22.4	12	0.85
11	CSSME70	4.33	10.7	27.7	550	261	4	0.58	0.574	0.498	0	47.9	9	0.78
11	SRCRSP6	10.56	5.8	38.4	537	394	4	0.68	0.663	0.612	0	59.9	9	0.73
11	FCB193	7.63	2.3	44.2	564	146	3	0.229	0.226	0.203	0	65.4	12	0.41
11	THRA	12.8	8.9	46.5	545	258	3	0.499	0.497	0.387	0.0138	67.2	8	0.73
11	EPCDV23	61.23	3	55.4	546	371	7	0.689	0.665	0.624	0	79.9	10	NA
11	MCM120	27.89	12.2	58.4	550	490	8	0.773	0.76	0.725	0	87.7	16	0.85
11	ETH3	3.57	20.6	70.6	536	319	3	0.565	0.557	0.491	0	99.4	5	NA
11	CSSM08			91.2	541	229	4	0.447	0.43	0.404	0	112.4	5	0.53
12	HUJ614	1.99	10.5	0	543	65	2	0.145	0.141	0.131	0	7.6	9	0.72
12	TGLA53	15.23	9.9	10.5	400	308	7	0.64	0.651	0.621	0.039	39.3	8	NA
12	BM4025	5.29	22.4	20.4	445	319	5	0.712	0.714	0.659	0.0074	24	9	NA
12	CSSM03	12.42	19.8	42.8	535	229	3	0.393	0.396	0.36	0	54.5	10	NA
12	MCMA52	3.5	27.4	62.6	536	387	5	0.674	0.678	0.624	0.0069	68.8	10	0.78
12	INRA35			90	523	212	5	0.512	0.477	0.436	0	92.1	9	0.68
13	BMC1222	19.35	15.5	0	508	416	5	0.734	0.706	0.67	0	12.3	15	0.82
13	SCYAMS	14.79	20.4	15.5	535	407	5	0.693	0.683	0.642	0.0197	37.4	20	0.89
13	MCM152	23.42	10.9	35.9	527	356	4	0.651	0.635	0.56	0.0087	52.1	10	0.79
13	HUJ616	13.51	10.3	46.8	546	365	5	0.615	0.609	0.538	0.0092	65	15	NA
13	URB058	4.59	17.5	57.1	548	254	4	0.5	0.508	0.387	0.0261	74.4	13	0.78
13	CTSBJ12	5.63	8.9	74.6	552	154	4	0.252	0.249	0.222	0	98	9	0.77
13	ADA	11.85	8.1	83.5	456	217	2	0.382	0.38	0.307	0	NA	NA	NA
13	MMP9	8.94	11.4	91.6	542	359	5	0.646	0.635	0.585	0	115.4	9	0.79
13	MAF18			103	335	171	3	0.573	0.559	0.488	0.019	125.8	5	0.41
14	TGLA357	41.95	5.1	0	525	315	6	0.651	0.657	0.606	0	11.7	8	0.8
14	INRA38	15.8	11	5.1	536	398	6	0.668	0.69	0.643	0	17.6	13	0.82
14	CSRD70	25.28	7.2	16.1	530	358	5	0.675	0.656	0.608	0	25.5	14	0.78
14	BMS2213	27.13	5.9	23.3	530	325	4	0.587	0.59	0.52	0	33.8	10	0.82
14	LS29	17.05	7.2	29.2	536	367	6	0.631	0.614	0.575	0	46.5	14	0.84
14	MCM133	11.53	11.2	36.4	532	185	3	0.361	0.362	0.31	0.0272	56.8	9	0.75

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	N[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l]		
												Pos[d]	No. all[g]	PIC[j]
14	CSRD32	8.34	23	47.6	527	293	4	0.545	0.511	0.459	0	64.6	15	0.82
14	LS30	21.14	6.5	70.6	532	308	4	0.609	0.635	0.568	0	94.4	11	0.79
14	RM128	36.86	2.2	77.1	556	303	4	0.482	0.473	0.436	0	104.6	11	0.81
14	MCMA19			79.3	519	189	3	0.374	0.354	0.326	0	109.3	6	NA
15	MCMA16	15.86	12.9	0	551	372	4	0.633	0.631	0.575	0.0081	0	9	0.63
15	BR3510	3.04	12.9	12.9	515	289	5	0.561	0.542	0.502	0	19.3	8	NA
15	BMS1004	6.33	0.98	25.8	532	99	2	0.195	0.189	0.171	0	27.2	13	NA
15	ADCYC	21.26	17	26.78	539	379	5	0.651	0.65	0.585	0	35.3	7	0.74
15	JAB1	23.14	1.2	43.78	533	388	8	0.683	0.703	0.655	0.0061	46.5	19	0.82
15	MAF65	0	29.8	44.98	328	172	4	0.518	0.512	0.453	0	47	8	0.62
15	HAEM	2.29	9.8	74.78	216	57	2	0.528	0.501	0.375	0	NA	NA	NA
15	POTCHA	30.64	5.3	84.58	510	413	5	0.716	0.717	0.668	0	85.1	11	NA
15	BMS1660	13.4	10	89.88	532	282	4	0.515	0.484	0.442	0	96.5	10	0.77
15	BMS2076	4.04	19.1	99.88	531	301	3	0.605	0.606	0.537	0	105.4	14	0.83
15	MCM105			118.98	534	500	8	0.785	0.784	0.75	0.0044	123.8	10	0.81
16	RM106	12.64	0	0	327	130	4	0.474	0.48	0.447	0	3.8	10	NA
16	BM1225	0	16.5	0	543	133	3	0.274	0.267	0.249	0	13.2	9	0.74
16	TGLA126	6.98	18.5	16.5	525	212	4	0.371	0.383	0.342	0.0247	34.3	18	0.84
16	AGLA29	30.04	6.7	35	502	308	6	0.631	0.637	0.581	0.0083	46.9	16	NA
16	CSRD69	6.28	3.7	41.7	519	361	6	0.667	0.681	0.636	0.0073	55.3	10	0.67
16	MCM506A	4.85	6.6	45.4	533	73	3	0.126	0.12	0.116	0	63.2	17	0.75
16	SRCRS27	4.58	5.2	52	509	385	6	0.68	0.688	0.631	0	69.4	8	0.72
16	MCM150			57.2	544	42	2	0.131	0.147	0.136	0	83.9	9	0.68
17	MCM4	0	16.9	0	550	404	5	0.596	0.662	0.616	0	0	13	0.85
17	VH98	9.19	6	16.9	553	71	4	0.179	0.178	0.172	0	19.7	9	0.67
17	CP49	22.83	12.6	22.9	543	449	7	0.703	0.687	0.649	0.0063	28.5	7	0.76
17	BMS2780	3.23	8.8	35.5	517	289	4	0.509	0.516	0.439	0.0147	38.4	8	0.75
17	FCB48	7.12	7.4	44.3	302	109	4	0.45	0.457	0.405	0.1119	42.5	11	0.76
17	MAF209	5.31	34	51.7	402	332	7	0.756	0.738	0.692	0.0515	48	8	0.79
17	MCMA20			85.7	507	489	6	0.984	0.806	0.777	0	89.2	11	0.73
18	BM3413	6.73	15.5	0	555	237	4	0.373	0.378	0.327	0	22.4	8	0.7
18	VH54	49.16	1.5	15.5	554	320	4	0.605	0.625	0.555	0	41.9	7	0.7
18	BP33	25.55	13.9	17	549	438	6	0.719	0.715	0.675	0	43.4	12	0.85
18	UWCA4	14.29	18.5	30.9	551	401	4	0.641	0.634	0.584	0	56.1	6	0.72
18	BMC5221	31.87	1.4	49.4	551	409	5	0.713	0.715	0.666	0	77	10	NA
18	HH47	2.5	16	50.8	379	205	5	0.623	0.658	0.607	0.0108	77	10	0.77
18	ILSTS54	0	16.8	66.8	552	89	2	0.239	0.248	0.217	0.1071	91.7	3	NA
18	IDVGA30	0	29.5	83.6	533	150	2	0.257	0.344	0.285	0.3782	110.5	2	NA
18	CSAP28E			113.1	558	261	4	0.461	0.439	0.393	0.0169	121.6	5	0.62
19	PZ963	12.91	16.9	0	549	350	6	0.65	0.618	0.574	0	10.6	22	NA

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	N[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l] Pos[d]	IMF map[l] No. all[g]	IMF map[l] PIC[j]
19	AE119	3.98	21.6	16.9	523	298	3	0.535	0.513	0.459	0	27.7	8	0.76
19	CSSM41	9.33	0	38.5	551	151	2	0.385	0.39	0.314	0.0219	NA	NA	NA
19	BM3628	2.47	40.1	38.5	499	264	5	0.515	0.511	0.47	0	43.3	4	NA
19	FCB304			78.6	577	352	4	0.591	0.586	0.506	0	66	9	0.54
20	BM1815	6.78	24.8	0	544	255	3	0.504	0.521	0.405	0.0867	26.8	6	NA
20	OLADRB	48.16	0	24.8	529	478	8	0.79	0.819	0.794	0.0149	52.2	13	NA
20	OLADRBps	23.85	5.6	24.8	263	203	6	0.795	0.786	0.752	0.014	NA	NA	NA
20	OMHC1	4.58	22	30.4	294	222	5	0.568	0.597	0.564	0.0325	NA	NA	NA
20	BM1818	6.45	19.2	52.4	533	388	8	0.657	0.678	0.629	0.0848	64.9	10	NA
20	BM1905			71.6	562	320	2	0.528	0.488	0.369	0	77.8	2	NA
21	BMS1787	7	11	0	524	438	4	0.716	0.711	0.656	0	15.5	16	0.84
21	RM044	5.07	9.8	11	520	125	3	0.26	0.251	0.23	0	22.1	10	0.83
21	CSAP30E	4.62	23.1	20.8	540	269	3	0.48	0.508	0.401	0	29.1	15	0.79
21	MCM135	2.84	22.9	43.9	493	399	5	0.996	0.676	0.618	0.0093	46	13	0.85
21	BMC1206			66.8	538	212	2	0.522	0.499	0.374	0	58.1	6	0.67
22	BMS907	48.34	10.1	0	498	480	7	0.815	0.838	0.817	0.0083	13.8	12	0.83
22	HEL11	24.14	9.1	10.1	531	509	8	0.825	0.849	0.83	0.0154	30	17	NA
22	BM1314	34.63	4.6	19.2	333	288	8	0.802	0.81	0.784	0	34.5	6	NA
22	INRA81	29.94	12.9	23.8	504	420	7	0.734	0.772	0.733	0	35.7	22	0.85
22	BM4505	32.57	11.9	36.7	518	402	7	0.707	0.734	0.69	0.0058	43.5	11	NA
22	BMS882	5.18	27.4	48.6	545	485	5	0.734	0.734	0.691	0	59.7	7	0.79
22	MCM373			76	531	231	6	0.405	0.411	0.393	0.021	82.9	13	0.84
23	BL6	8.13	20.8	0	520	310	7	0.619	0.636	0.576	0.0091	15.7	12	NA
23	CSRD148	12.28	9.2	20.8	502	319	5	0.697	0.673	0.616	0	33.2	15	0.82
23	BMS2270	9.05	11.7	30	557	224	5	0.361	0.361	0.339	0	37.7	8	0.79
23	AGLA269	13.22	15.2	41.7	508	333	5	0.646	0.621	0.561	0	49.1	18	NA
23	MAF35	10.49	10.5	56.9	583	369	3	0.568	0.565	0.473	0	59.2	5	0.61
23	MCM136	0	37.6	67.4	552	438	5	0.737	0.743	0.694	0	67.6	8	0.83
23	URB031			105	516	148	4	0.347	0.369	0.347	0.0567	97	7	0.62
24	EPCDV03	45.75	3.6	0	529	344	4	0.59	0.565	0.508	0.0108	28.9	8	NA
24	BP28	21.85	14.8	3.6	540	400	4	0.67	0.664	0.611	0.0143	38.5	18	0.92
24	FIBROSN	1.9	31.4	18.4	549	421	4	0.727	0.724	0.674	0.0057	48.9	12	0.82
24	EPCD152			49.8	525	439	5	0.985	0.656	0.587	0	83.2	6	NA
25	MCMA7	12.21	8.4	0	528	224	3	0.422	0.391	0.337	0	31	12	0.9
25	PERF3-2	16.39	6.8	8.4	525	382	5	0.714	0.671	0.615	0.0077	40	15	0.86
25	VH72	3.18	21.1	15.2	522	171	4	0.276	0.26	0.247	0	45.6	7	0.76
25	AE54	3.97	15.1	36.3	402	231	5	0.617	0.632	0.567	0.0442	64	9	0.82
25	RBP3			51.4	332	200	3	0.642	0.608	0.536	0	69.9	4	0.61
26	BMS629	19.53	8	0	527	309	3	0.575	0.579	0.49	0	6.9	9	NA
26	BM6526	26.85	2.3	8	541	336	4	0.534	0.536	0.48	0.0121	16.4	9	NA

(*continued*)

APPENDIX

(Continued)

Chromosome[a]	Marker	Two-point LOD[b]	Inter[c] (cM)	Pos[d] (cM)	N[e]	InfMei[f]	No. all[g]	H(O)[h]	H(E)[i]	PIC[j]	Est err rate[k]	IMF map[l] Pos[d]	IMF map[l] No. all[g]	IMF map[l] PIC[j]
26	LS41	7.1	18.1	10.3	535	310	4	0.624	0.617	0.542	0.0091	20.6	12	0.77
26	CSRD163	17.18	16.9	28.4	524	270	3	0.578	0.614	0.531	0.0097	39	8	NA
26	JMP23	20.7	5.3	45.3	531	443	8	0.727	0.731	0.681	0.0058	53.5	13	NA
26	JMP58	9.71	8.9	50.6	401	222	5	0.589	0.596	0.522	0.0263	51.4	9	0.67
26	POLBF17	11.2	14.8	59.5	497	328	7	0.678	0.669	0.641	0	61.5	11	0.82
26	BM203			74.3	335	278	9	0.773	0.79	0.759	0.0424	71.1	10	NA
X	MCM158	4.27	34.6	0	528	397	7	0.714	0.648	0.599	0	0.8	12	0.88
X	MAF45	18.21	24.7	34.6	709	623	6	0.779	0.742	0.701	0	31.2	12	0.84
X	ILSTS17	66.49	5.1	59.3	530	543	3	0.577	0.47	0.41	0	66.8	11	0.79
X	CP131	72.24	1.8	64.4	524	537	7	0.517	0.525	0.494	0	80.5	9	0.83
X	MCM25			66.2	528	464	5	0.729	0.628	0.579	0	90.8	15	NA
	Average	14.58	15.0		510	310	4.58	0.58	0.58	0.52	0.01		10.06	0.75
	SD	14.88	9.9		73	113	1.58	0.17	0.16	0.16	0.03		3.49	0.11

[a] Chromosome number.
[b] LOD score for linkage between adjacent markers.
[c] Intermarker spacing.
[d] Chromosomal position.
[e] Number of sheep genotyped.
[f] Number of informative meioses detected at the marker locus.
[g] Number of alleles.
[h] Observed marker heterozygosity.
[i] Estimated marker heterozygosity.
[j] Polymorphism information content.
[k] Estimated error rate from mother–offspring pairs using CERVUS.
[l] IMF map characteristics (Australian Sheep Gene Mapping website at http://rubens.its.unimelb.edu.au/~jillm/jill.htm).

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

International Journal for Parasitology xxx (2006) xxx–xxx

www.elsevier.com/locate/ijpara

Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*)

Dario Beraldi [a,*], Allan F. McRae [a,b], Jacob Gratten [c], Jill G. Pilkington [a], Jon Slate [c], Peter M. Visscher [a,b], Josephine M. Pemberton [a]

[a] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK*
[b] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane 4029, Australia*
[c] *Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK*

Received 27 July 2006; received in revised form 4 September 2006; accepted 15 September 2006

Abstract

A genome-wide scan was performed to detect quantitative trait loci (QTL) for resistance to gastrointestinal parasites and ectoparasitic keds segregating in the free-living Soay sheep population on St. Kilda (UK). The mapping panel consisted of a single pedigree of 882 individuals of which 588 were genotyped. The Soay linkage map used for the scans comprised 251 markers covering the whole genome at average spacing of 15 cM. The traits here investigated were the strongyle faecal egg count (FEC), the coccidia faecal oocyst count (FOC) and a count of keds (*Melophagus ovinus*). QTL mapping was performed by means of variance component analysis so that the genetic parameters of the study traits were also estimated and compared with previous studies in Soay and domestic sheep. Strongyle FEC and coccidia FOC showed moderate heritability ($h^2 = 0.26$ and 0.22, respectively) in lambs but low heritability in adults ($h^2 < 0.10$). Ked count appeared to have very low h^2 in both lambs and adults. Genome scans were performed for the traits with moderate heritability and two genomic regions reached the level of suggestive linkage for coccidia FOC in lambs (logarithm of the odds = 2.68 and 2.21 on chromosomes 3 and X, respectively). We believe this is the first study to report a QTL search for parasite resistance in a free-living animal population and therefore may represent a useful reference for similar studies aimed at understanding the genetics of host-parasite co-evolution in the wild.

Keywords: QTL mapping; Soay sheep; Parasitic nematodes; Variance components; Natural population

1. Introduction

The antagonism between host and parasite is thought to be a major force in ecology and evolution due to its potential to generate and maintain genetic variation. Parasites are often characterized by high potential for diversification due to their high speed of speciation (Dykhuizen, 1998), whereas the hosts they colonize constitute a rapidly changing environment (Huyse et al., 2005). As a consequence, the host-parasite relationship generates continuously evolving host and parasite lineages (Nadler, 1995). In principle, this continuous battle can maintain genetic diversity in the antagonistic populations provided that a specificity between host and parasite genotypes is present (Haldane, 1949). Parasite resistance is likely to be controlled by several loci and therefore it may receive a strong mutational input which generates genetic variation (Houle et al., 1996). Host-parasite co-evolution may maintain genetic variation if the additive genetic value of a host genotype changes when parasites evolve as a response to the selection induced by the host (Haldane, 1949). Antagonistic pleiotropy may result in maintenance of genetic variation if the same genotype is positively selected for one fitness-related trait but negatively selected for another fitness-related trait (Roff and Mousseau, 1987). In the case of parasite resistance, this last hypothesis is suggested by the finding that

* Corresponding author. Tel.: +44 131 6513612; fax: +44 131 6506564.
E-mail address: dario.beraldi@ed.ac.uk (D. Beraldi).

0020-7519/$30.00
doi:10.1016/j.ijpara.2006.09.007

sheep that are genetically resistant to intercellular infections may be more susceptible to infection from intracellular pathogens (Gill et al., 2000).

In sheep and other domestic ruminants, gastrointestinal nematodes are one of the most important classes of parasite. Intensive effort, therefore, has been invested in understanding, and exploiting through breeding programs, the genetic basis of parasite resistance and host-parasite co-evolution (Kaplan, 2004). Parasite resistance is complex in nature, having polygenic and environmental components (Stear et al., 1997; Bishop and Stear, 2003). Resistance to infection by gastrointestinal nematodes has moderate heritability in domestic sheep ranging from 0.13 (McEwan et al., 1992) to 0.53 (Baker et al., 1991) and resistant or susceptible lines have been selected in various countries (Dominik, 2005).

Quantitative trait loci (QTL) mapping can help to dissect the complexity of parasite resistance by identifying candidate genomic regions affecting the trait variation. To this end, different linkage mapping projects have been undertaken to find QTL for parasite resistance (Dominik, 2005). A genome scan was performed by Beh et al. (2002) using lines of sheep diverging for parasite resistance. Different regions were detected as likely to carry genes for resistance although no region was statistically significant after correcting for multiple tests. Davies et al. (2006) genotyped naturally infected lambs to scan regions previously identified as candidates for either genes for resistance or genes for other economical traits to determine whether these candidate regions could be confirmed in an independent dataset. Evidence of linkage was found on chromosomes 2, 13, 14 and 20. Recently, a genome scan performed by Crawford et al. (2006) using divergent lines and naturally infected animals detected a significant QTL on chromosome 8. In general, such selected populations have the advantage of high quality pedigrees and phenotypic data in terms of sample size and accuracy. In order to increase the power of analysis, the populations used for mapping purposes are usually grown under controlled and uniform conditions designed to maximize the genetic contribution to the phenotype (Lymbery, 1996; Lynch and Walsh, 1998). Such experimental designs, however, may not accurately reflect the interactions between genes and environment that occur in natural populations (Erickson et al., 2004; Slate, 2005) and that could contribute to host-parasite co-evolution and host population dynamics (Gulland and Fox, 1992; Gulland et al., 1993; Hudson et al., 1998).

From an evolutionary perspective, it is of interest to know whether major genes for parasite resistance explain observed variation in natural populations in situ. Perhaps such genes can only be detected under highly controlled environmental conditions when genetic variation and statistical power are maximised. In free-living populations, environmental noise and interactions between genetic variation and environmental variation may mask the effects of individual genes (Lynch and Walsh, 1998). In this paper we present a QTL analysis of parasite resistance in a free-living sheep population.

The free-living Soay sheep population on Hirta, St. Kilda, UK, is the subject of a long term project aimed at addressing a wide range of ecological and evolutionary issues (extensively documented in Clutton-Brock and Pemberton, 2004) including the genetics and evolution of parasite resistance in the wild. The Soay sheep is naturally parasitized by several gastrointestinal nematode species (Wilson et al., 2004; Wimmer et al., 2004; Craig et al., 2006), the most prevalent and abundant being strongyles (of which the predominant species are *Teladorsagia circumcincta*, *Trichostrongylus axei*, *Trichostrongylus vitrinus*; see Craig et al., 2006). Different species of protozoans also infect the intestinal tract of Soay sheep; these belong mainly to the genus *Eimeria* but *Cryptosporidium parvum* and *Giardia duodenalis* also occur (Wilson et al., 2004; Craig et al., in press). Keds (*Melophagus ovinus*) also parasitize the Soay sheep, living in the wool and feeding on blood, causing anaemia and irritation (Wilson et al., 2004).

The evolution of parasite resistance in Soay sheep has previously been addressed using two different population genetics strategies, the first being a quantitative genetics approach. Parasite resistance, measured as strongyle faecal egg count (FEC), is under directional selection (Coltman et al., 1999). In addition, there is a positive genetic correlation between body traits and resistance to strongyles (Coltman et al., 2001a) so that resistant sheep also experience better growth. Because body size and parasite resistance are under directional selection, it is expected that in this population the allelic variants associated with small body size and/or low parasite resistance will be eliminated by selection and additive genetic (heritable) variation will be reduced to near zero (Fisher, 1958; Endler, 1986). However, parasite resistance in Soay sheep has low but not null heritable variation – previous population-wide estimates based on an animal model found a heritability for FEC in summer of 0.11 ± 0.02 in males and 0.13 ± 0.01 in females (Coltman et al., 2001a).

In a second approach, previous studies in Soay sheep have examined a number of candidate loci for parasite resistance in simple association studies. Three loci appear to be associated with parasite resistance: the interferon gamma gene (IFNG) on chromosome 3 (Coltman et al., 2001b); the major histocompatibility complex (MHC) on chromosome 20 (Paterson, 1998; Paterson et al., 1998); and the adenosine deaminase gene (ADA) on chromosome 13 (Gulland et al., 1993). It is of great interested to know whether these association studies can be supported in a more rigorous QTL search.

Here, we make use of a previously established mapping pedigree and linkage map (Beraldi et al., 2006; Beraldi et al., unpublished data) and phenotypic data for three classes of parasites (gastrointestinal nematodes, coccidia and keds), to ask: (i) whether we can detect heritable variation for resistance; (ii) whether we can detect QTL for resistance; and (iii) whether any QTL found coincide with previous domestic sheep or Soay sheep studies.

Please cite this article in press as: Beraldi, D. et al., Quantitative trait loci (QTL) mapping of resistance to strongyles ..., Int. J. Parasitol. (2006), doi:10.1016/j.ijpara.2006.09.007

2. Materials and methods

2.1. Study population

The Soay sheep on the islands of Soay and Hirta (St. Kilda archipelago, North West Scotland, UK, 57°49′N, 08°34′W) are feral populations of a breed regarded as the most primitive in Europe (Campbell, 1974; Doney et al., 1974); nowadays, the sheep population of Hirta varies between 600 and 2000 individuals. Since 1985 regular expeditions have been sent to St. Kilda to monitor the population dynamics and to record the life histories of individuals living in Village Bay, Hirta (Clutton-Brock et al., 2004). No predators are present on St. Kilda. All animal handling was undertaken under the appropriate UK Home Office licences.

2.2. Mapping pedigree and linkage map

The whole Soay sheep pedigree file numbers more than 3900 animals. Within this pedigree maternal links were assigned through observation of the animals in the field, whereas paternal links were inferred through molecular analysis (Overall et al., 2005). From the total pedigree, a panel of 588 animals was genotyped at 247 microsatellite and four isoenzyme markers. This subset comprised all the half-sibships with 10 or more individuals and their common parents. The ancestors of the genotyped individuals and the animals linking different sibships ($n = 294$) were not genotyped, but they were included in the mapping pedigree to improve the estimates of kinship and the identity by descent (IBD) coefficients in the variance component analysis. A more thorough description of the mapping pedigree and selection criteria is included in Beraldi et al. (2006). The Soay sheep map covers approximately 90% of the genome with an average inter-marker spacing of 15 cM. Further details of the map characteristics and the technical procedures can be found in Beraldi et al. (2006).

2.3. Phenotypic dataset and measures of parasitism

Phenotypic records of the animals in the mapping pedigree were retrieved from the Soay sheep database. The data analyzed in this study were collected between 1988 and 2005 from animals born between 1979 and 2002.

In the present study, the quantification of sheep resistance to gastrointestinal parasites was based on the indirect measures of strongyle FEC and coccidia faecal oocyst count (FOC). The direct count of parasites would involve the sacrifice of animals and post-mortem examination: this alternative is not feasible because the Soay sheep are protected and the sacrifice of animals would be in conflict with the study of the Soay sheep as a free-living population. However, previous work has shown a correlation between FEC and burden in island Soay populations (Wilson et al., 2004). Strongyle FEC and coccidia FOC were determined as the number of parasite eggs (FEC) or oocysts (FOC) per gram (wet weight) of faeces using a modification of the McMaster technique (MAFF, 1986). Other distinctive helminth species (*Nematodirus* spp., *Moniezia expansa*, *Capillaria longipes* and *Trichuris ovis*) are routinely classified and quantified in Soays but were not abundant enough for analysis. A few hosts that had previously been treated with either anthelminthics or hormones for experimental purposes were excluded from analysis. The count of keds was the total number of keds observed during a 1 min search of the wool on a sheep's belly. The raw data (strongyle FEC, coccidia FOC and ked count) were transformed into the natural logarithm to achieve a distribution closer to normality (all the measurements were increased by one unit before transformation, i.e. Ln(trait + 1), so that zero values remained unchanged after transformation). The genetic and environmental sources of variation of parasite resistance are expected to change with the age of the animals and time of year (Bishop et al., 1996; Coltman et al., 2001b). Therefore, only the samples collected in the August catch up, when most of the data are collected, were included in the analyses. In addition, each parasitic group was analyzed separately in lambs (4-month-old animals) and adults (animals older than 4 months). Sample sizes and summary statistics for each trait are reported in Table 1.

2.4. Definition of fixed effects

Fixed effects influencing the study traits were fitted in the variance component models. In order to facilitate comparisons with previous studies in Soay sheep, the fixed effects fitted for strongyle FEC were the same as those fitted by Coltman et al. (2001b). For consistency, coccidia FOC was also analyzed with the same model. A general linear model analysis implemented in Minitab 14.1 (Minitab Inc.) was applied to determine the amount of variation explained by each fixed effect (Table 2).

2.5. Estimation of variance components

Under the null hypothesis of no segregating QTL, the additive genetic variation of a trait is supposed to be composed by many genes of small effect scattered across the genome. The trait can be modelled as a combination of fixed and random effects (Lynch and Walsh, 1998; Williams and Blangero, 1999):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of records on individuals; $\boldsymbol{\beta}$ is a vector of fixed effects; $\mathbf{a}$ is a vector of additive genetic effects (or breeding values) estimated on the basis of the coefficient of co-ancestry between any pair of individuals in the pedigree; $\mathbf{e}$ is a vector of residual effects. $\mathbf{X}$ and $\mathbf{Z}$ are design matrices relating records to the appropriate fixed or random effects.

Heritability (h^2), permanent environment effect (c^2) and residual effect (e^2) were calculated as the ratio of the rele-

Please cite this article in press as: Beraldi, D. et al., Quantitative trait loci (QTL) mapping of resistance to strongyles ..., Int. J. Parasitol. (2006), doi:10.1016/j.ijpara.2006.09.007

Table 1
Characteristics and estimated variance components of the study traits

Trait	Dataset	No. records	No. animals (genotyped)	Mean (SE)[a]	V_A[b] (SE)[a]	CV_A[c] (%)	h^2[d] (SE)[a]	V_C[e] (SE)[a]	CV_C[f] (%)	c^2[g] (SE)[a]	V_E[h] (SE)[a]	CV_E[i] (%)	e^2[j] (SE)[a]
Strongyles	Lambs	383	381 (307)	5.376 (2.150)	0.944 (0.472)	18.08	0.26 (0.12)	NS	—	—	2.626 (0.441)	30.14	0.74 (0.12)
Coccidia	Lambs	230	228 (204)	7.979 (1.228)	0.254 (0.254)	6.32	0.22 (0.21)	NS	—	—	0.915 (0.243)	11.99	0.78 (0.21)
Keds	Lambs	376	374 (310)	1.221 (0.804)	0.022 (0.054)	12.06	0.04 (0.09)	NS	—	—	0.563 (0.067)	61.45	0.96 (0.09)
Strongyles	Adults	962	345 (192)	3.073 (2.749)	0	—	—	0.789 (0.236)	28.90	0.13 (0.04)	5.479 (0.311)	76.17	0.87 (0.04)
Coccidia	Adults	694	240 (155)	5.224 (2.517)	0.310 (0.183)	10.66	0.06 (0.03)	NS	—	—	4.890 (0.305)	42.33	0.94 (0.03)
Keds	Adults	1303	396 (229)	0.123 (0.312)	0.002 (0.002)	40.97	0.03 (0.02)	NS	—	—	0.089 (0.004)	242.50	0.97 (0.02)

[a] Standard error.
[b] Additive genetic variance.
[c] Coefficient of additive genetic variation.
[d] Heritability.
[e] Permanent environmental variance.
[f] Coefficient of permanent environmental variation.
[g] Permanent environmental effect.
[h] Residual variance.
[i] Coefficient of residual variation.
[j] Residual effect.

vant variance component (V_A, additive genetic variance; V_C, permanent environmental variance; V_E, residual variance) to total phenotypic variance (V_P), i.e. $h^2 = V_A/V_P$; $c^2 = V_C/V_P$; $e^2 = V_E/V_P$.

The coefficients of variation were calculated as:

$$CV_i = 100V_i^{1/2}/\bar{x},$$

where the subscript i stands for the additive genetic (A), permanent environment (C) and residual components (E) and $\bar{x}$ is the trait mean.

Variance components were estimated by the restricted maximum likelihood procedure (Lynch and Walsh, 1998) implemented in the software package ASReml (Gilmour et al., 2002).

2.6. QTL mapping

To map putative segregating QTL, an IBD (identity by descent) matrix estimated at any given map position was fitted in the model described above as an additional random effect (George et al., 2000):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Zq} + \mathbf{e}$$

where $\mathbf{q}$ is a vector of additive QTL effect. IBD sharing statistics were estimated using pedigree relationships, marker data and map distances described above and in Beraldi et al. (2006). For an initial scan, IBD matrices and variance components were estimated every 10 cM. Putative QTL regions, i.e. those reaching a logarithm of the odds (LOD) score of at least 1, were then scanned every 1 cM. The calculation of the IBD matrices was performed by a Markov chain Monte Carlo (MCMC) which allows the handling of very large and complex pedigrees. This process was implemented in the program Loki (Heath, 1997). LOD scores were calculated as the difference in log-likelihood between QTL and polygenic model according to the equation:

$$LOD = (L_1 - L_0)/\ln(10)$$

where L_1 is the natural log-likelihood of the QTL model and L_0 the natural log-likelihood of the polygenic model.

Genome-wide suggestive and significant thresholds were obtained by solving Equation 1 of Lander and Kruglyak (1995) assuming a map length of 33.5 Morgans spanning 27 chromosomes. Solutions were 1.9 and 3.3, respectively. The genome-wide significance (LOD = 3.3) corresponds to the probability of finding a false positive every 20 genome scans; the suggestive significance (LOD = 1.9) corresponds to the probability of finding a false positive once per genome scan (Lander and Kruglyak, 1995). Here, all the LOD scores exceeding the arbitrary threshold of 1 are reported. For LOD scores above the suggestive threshold, support intervals for the presence of a putative QTL were defined by the map range within a one-LOD score drop from the peak value; which is equivalent to approximately 95% confidence (Lander and Botstein, 1989).

Please cite this article in press as: Beraldi, D. et al., Quantitative trait loci (QTL) mapping of resistance to strongyles ..., Int. J. Parasitol. (2006), doi:10.1016/j.ijpara.2006.09.007

Table 2
Fixed effects for the study traits fitted in the polygenic and quantitative trail loci models

Trait	Dataset	Sex	Litter size	Birth year	Collection year	Weight	Collection age	Total deviance explained (df)
Strongyles	Lambs	4.2 (1)	0.3 (2)[a]	—	17.4 (14)	3.5 (1)	NF	25.5 (18)
Coccidia	Lambs	0.7 (1)[a]	~0 (2)[a]	—	20.8 (9)	5.6 (1)	NF	27.4 (13)
Ked count	Lambs	5.8 (1)	3.1 (2)	—	4.1 (14)[a]	NF	NF	13.0 (17)
Strongyles	Adults	13.2 (1)	NF	NF	4.0 (17)	0.4 (1)	NF	21.2 (22)
Coccidia	Adults	1.4 (1)	NF	NF	13.4 (12)	1.2 (1)	2.1 (1)	18.2 (15)
Ked count	Adults	NF	NF	4.1 (22)	5.8 (16)	NF	NF	9.9 (38)

The deviance explained (in percentage) and the degrees of freedom used by each term (in brackets) are reported.
NF, Not fitted.
[a] Effect non-significant ($P > 0.05$) but fitted for consistency with previous analyses.

3. Results and discussion

Six (non-independent) traits reflecting the resistance of Soay sheep to gastrointestinal strongyles, coccidia and keds at ages 4 months and 16 months or older were modelled.

3.1. Variance component analysis

Results of the variance component analysis under the polygenic model are presented in Table 1. In lambs, the additive genetic component of strongyle FEC and coccidia FOC accounted for a moderately high proportion of the phenotypic variation although the estimates were not very precise due to the large standard deviations. The heritabilities of strongyle FEC and coccidia FOC in lambs were similar, being 0.26 ± 0.12 and 0.22 ± 0.21, respectively. In adults, no genetic variation was detected for strongyle FEC and very low heritability was detected for coccidia FOC ($h^2 = 0.06 \pm 0.03$). The estimates of heritability reported by Coltman et al. (2001a) for strongyles FEC in Soay sheep were between 0.11 and 0.14. The inconsistency between results from Coltman et al. and the present study could be explained by differences in the pedigree and data selection. In particular, the estimates of Coltman et al. (2001a) are based on animals of any age whereas in this study we differentiated between lambs (4-months-old animals) and adults (animals older than 4 months). Also, higher estimates of heritability in this study may be explained by more reliable inference of parentage. The genotyping of more than 200 markers in the genome scan allowed the detection of pedigree errors that can downwardly bias the estimate of genetic parameters (Charmantier and Reale, 2005). With respect to domestic sheep, the estimates of FEC heritability in Soay sheep reported here do not particularly differ from farmed or experimental populations. However, as heritability is a property of the population and not the species, care should be taken in comparing Soay and domestic sheep because of the differences in life history and environment between a free-living population and managed, selected flocks.

With respect to the coefficients of variation, the CV_A of strongyle FEC in lambs was about three times the CV_A of coccidia FOC in lambs (18.08 and 6.32, respectively) although the heritability of the two traits was similar (0.26 and 0.22, respectively). This suggests that there is greater genetic variation responsible for strongyle FEC than coccidia FOC, but also that the phenotypic variation of strongyle FEC is higher than that of coccidia FOC.

The adult datasets of strongyle FEC, coccidia FOC and ked count in both lambs and adults showed little or zero heritable variation (Table 1). However, the CV_A in adult coccidea FOC (10.66) is higher than in lambs (6.32). This suggests the genetic variation in adults is overwhelmed by environmental variation so that the genetic component makes little contribution to phenotype. The same speculation could be applied to strongyle FEC and ked count but in this case the fact that the means and variances are similar and close to zero make the coefficient of variation difficult to interpret due to its mathematical properties.

3.2. Variance component QTL mapping

Genome scans were performed for strongyle FEC and coccidia FOC in lambs, the two traits with moderate heritability. The LOD score profiles for these traits are shown in Fig. 1 and characteristics of LOD scores higher than 1 are listed in Table 3.

Three LOD scores above 1 but below the suggestive threshold were detected for strongyle FEC (Table 3). These were located on chromosomes 6 (LOD = 1.58), 12 (LOD = 1.49) and 1 (LOD = 1.43). Beh et al. (2002) detected a suggestive QTL for resistance to *Trichostrongylus colubriformis* in 20-week-old sheep on chromosome 6 and a pointwise significant peak (significant at $P < 0.05$ but unadjusted for multiple tests) for 27-week old animals. This group also identified one region on each of chromosomes 1 and 12 reaching the pointwise significance in 27-week-old animals. The scan published in that study does not report the position of the LOD peaks so that it is not possible to determine whether their peaks correspond to those presented here. To the best of our knowledge no study, other then that mentioned above, detected QTL for parasite resistance in the regions reported here.

The scan for coccidia FOC in lambs produced two LOD scores exceeding the suggestive threshold (chromosome 3 with LOD = 2.68 and support interval of approximately 30 cM, and chromosome X with LOD = 2.21 and support interval of approximately 17 cM; Table 3 and Fig. 2), and

Please cite this article in press as: Beraldi, D. et al., Quantitative trait loci (QTL) mapping of resistance to strongyles ..., Int. J. Parasitol. (2006), doi:10.1016/j.ijpara.2006.09.007

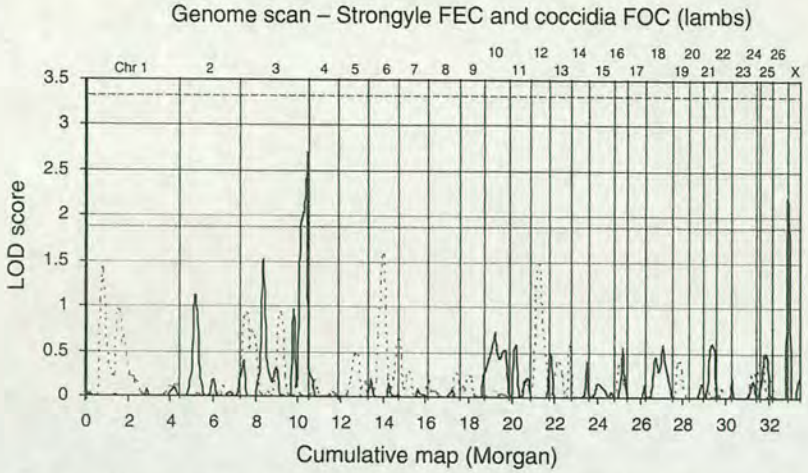

Fig. 1. Whole genome scans of strongyle faecal egg count (FEC; dashed line) and coccidia faecal oocyst count (FOC; continuous line) in lambs. Logarithm of the odds (LOD) score values (ordinate) were plotted against genetic position (abscissa, Morgan scale). Dotted horizontal lines show the genome-wide significance threshold (3.3); dashed lines are the suggestive significance threshold (1.9). Vertical lines mark the chromosome boundaries and chromosome names are displayed at the top.

Table 3
Logarithm of the odds (LOD) scores higher than 1 detected for strongyle faecal egg count and coccidia faecal oocyst count in lambs

Trait	Dataset	LOD	Chr.	Position (cM)	Flanking markers (cM)[a]	
Strongyles	Lambs	1.58	6	74	BMS360 (4)	McM140 (4)
		1.49	12	44	CSSM3 (1)	MCMA52 (19)
		1.43	1	79	BM6465 (7)	CSAP36E (1)
Coccidia	Lambs	2.68[b]	3	328	CSAP39E (17)	CSSME76 (0)
		2.21[b]	X	3	McM158 (3)	MAF45 (32)
		1.52	3	127	RM96 (6)	BM2818 (14)
		1.13	2	89	CSSM37 (7)	FCB128 (3)

[a] In parentheses the distance (cM) of the flanking markers from the quantitative trait loci peak.
[b] Suggestive linkage (LOD >1.9).

two LOD scores exceeding the value of 1 (chromosome 3 with LOD = 1.52 and chromosome 2 with LOD = 1.13; Table 3). The LOD score peak for coccidia FOC in lambs on chromosome X is in the vicinity of one of the telomeres. No previous studies have investigated chromosome X for parasite resistance QTL or genes. Davies et al. (2006) identified three regions on chromosome 3 likely to be linked to different traits related to parasite resistance (IgA activity, *Nematodirus* FEC in August, and strongyle FEC in October) and one region on chromosome 2 linked to *Nematodirus* FEC in September. Although the support intervals of the present study and that of Davies et al. (2006) overlap, the statistical significance of the results and the differences in the two approaches make it difficult at this stage to understand whether the two studies have identified the same regions. Other studies in domestic sheep did not identify QTL in the regions detected here.

Candidate regions identified in previous association studies in Soay sheep did not produce any evidence of linkage in our genome scans. It is known that the IFNG region on chromosome 3 is related to strongyle parasite resistance in domestic sheep (Paterson et al., 1999) and Soay lambs (Coltman et al., 2001b). The IFNG region, located at approximately 244 cM on the Soay map, did not produce any particular evidence of linkage and it is outside the support interval of the suggestive QTL identified for coccidia FOC. The effect of IFNG on strongyle FEC in Soay sheep was estimated by Coltman et al. (2001b) in a dataset larger than the one analyzed here and using a general linear model in which the two alleles of a microsatellite in the IFNG gene were fitted as a fixed factor (Coltman et al., 2001b). The association between FEC and microsatellite alleles was significant at $P = 0.047$ but the variation explained by the microsatellite was very low (0.5% of the total deviance). Consequently, it is not surprising that no QTL was detected in this region considering the differences in datasets and methods of analysis. Similarly, the ADA locus, mapped to chromosome 13 (Beraldi et al., 2006), and the MHC region on chromosome 20 did not produce any evidence of linkage. As with IFNG, the association between the MHC region and strongyle FEC was detected using a larger dataset with a generalized linear model. Failure to detect linkage could be due to insufficient power of the current sample size or lack of informative markers in the target regions although good marker coverage was achieved in the putative regions (Beraldi et al., 2006). However, associ-

Please cite this article in press as: Beraldi, D. et al., Quantitative trait loci (QTL) mapping of resistance to strongyles ..., Int. J. Parasitol. (2006), doi:10.1016/j.ijpara.2006.09.007