
THE UNIVERSITY
*of* EDINBURGH

# Unravelling major histocompatibility complex diversity in the Soay sheep of St Kilda

**Kara L. Dicks**


THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

The University of Edinburgh

2017

# Abstract

The major histocompatibility complex (MHC) is one of the most variable regions in the vertebrate genome. Many genes within the MHC play important roles in the development of an immune response, including the response to pathogens, by presenting pathogen fragments to T cells. Pathogen-mediated balancing selection is thought to be important in maintaining the high levels of allelic variation at these loci, though the precise mechanism remains unclear. The number of studies of MHC diversity in non-model organisms has increased dramatically in recent years as genotype data have become cheaper and easier to generate; however, key limitations in many such studies remain a lack of high quality MHC genotypes and associated phenotype data. Many studies focus on a single MHC locus, assuming that one locus will represent the full range of variation within each MHC haplotype. Alternatively, the products of different loci may co-amplify, preventing locus-specific genotypes and hence heterozygosity being accurately determined. Non-model systems are also often limited by small sample sizes and limited recording of associated host and pathogen measures, which, combined with high levels of allelic variation at MHC loci, can limit statistical power. Finally, few MHC studies control for the general effect of relatedness in explaining host traits before testing for MHC effects. With so many methodological impediments, it is challenging to identify robust associations between MHC variation and host phenotypes, such as parasite burden or fitness, and to draw conclusions about the mechanisms underpinning the maintenance of diversity at MHC loci.

In this thesis, I address these problems by developing a SNP-based haplotyping system for a population of unmanaged Soay sheep (*Ovis aries*) on Hirta, St. Kilda, for which data is available on pedigree, phenotypic traits and fitness and its components over a 30-year study period. The ovine MHC consists of four classes of loci, within which loci are tightly clustered and show reduced recombination rates compared to the genome average. Although the mammalian MHC is usually highly variable, one would expect that the number of haplotypes within an MHC class in an island population of sheep with no immigration to be limited. The class IIa region of the ovine MHC includes the classical class II loci which are typically thought to be involved in the presentation of

peptides derived from extracellular pathogens, including gastrointestinal helminths, in sheep and other mammals. In chapters 2 to 4, I describe the characterisation of class IIa haplotypic diversity in the Soay sheep using direct Sanger sequencing of PCR amplified fragments, which, in combination with cloning, revealed eight distinct haplotypes. With this knowledge of haplotypic diversity, and genotypes for a sample of Soay sheep typed on the Ovine Infinium HD chip (approximately 600K SNPs), I developed a panel of 13 SNPs which could be used to impute the class IIa haplotypes. This panel was genotyped by KASP (Kompetitive Allele Specific PCR) in 6034 samples and used to impute the class IIa haplotypes. After quality control measures, class IIa haplotypes were successfully imputed for 5349 individuals. Evidence of balancing selection was identified using the Ewens-Watterson test at different life history stages and within the standing population each year between 1985 and 2012, showing that allele frequencies were more even than would be expected under neutrality. However, there was no evidence of deviation from Hardy-Weinberg equilibrium identified at different life stages or in the standing population in any year.

In chapter 5, I investigate associations between the MHC class IIa haplotypes and individual-level data on host phenotypes – body weight, plasma immunoglobulin levels (measured as anti-*Teladorsagia circumcincta* third larval stage IgA, IgE and IgG) and strongyle faecal egg counts (FEC). Associations were tested within mixed effects models which were used to account for repeated measures and control for fixed effects known to affect the response variables, as well as within an animal model framework to account for relatedness between individuals. Haplotype heterozygosity was unrelated to any of the traits investigated, suggesting a general heterozygote advantage is unlikely to be operating within the Soay sheep. Six of the eight class IIa haplotypes were associated with multiple traits in different age-sex classes, although many of these associations were removed after inclusion within animal models. The evidence of balancing selection and associations between class IIa haplotypes and phenotypes related to health offers a promising glimpse into the evolutionary mechanisms which may be operating to maintain diversity within this region.

# Lay summary

The major histocompatibility complex (MHC) is one of the most variable regions in the vertebrate genome. There are many genes within the MHC, and these genes usually have many variants, known as alleles. A major question in evolutionary biology is why these genes are so variable and why the variation remains within the population. Why do the most beneficial alleles not spread throughout the population, replacing less beneficial alleles? The MHC genes encode MHC molecules which are located at the cell surfaces and are involved in alerting the immune system to invading pathogens. They do this by presenting small fragments of proteins that originate from the invading pathogens to a specific type of immune cell, T cells, which trigger an immune response. Different alleles at MHC genes generate different variants of MHC molecules, which are then able to each present protein fragments from a different range of pathogens.

Diversity at the MHC is thought to be maintained by balancing selection (the maintenance of multiple alleles in a population) driven by diversity within pathogen communities. Exactly how balancing selection maintains variation at the MHC has interested evolutionary biologists for many years, and yet studies are typically limited in their ability to answer this question. One way to address this question is to genotype the MHC in large numbers of individuals and compare them to measures of individual health and fitness. Whilst experimental systems are rarely able to replicate the levels of diversity which occur in the wild, wild systems are difficult to study because of that diversity. This creates technical challenges, in particular genotyping the multiple, highly variable genes in large numbers of individuals.

In this thesis, I make use of long-term data and genetic samples collected for the unmanaged population of Soay sheep that have been living feral on the islands of St. Kilda for thousands of years. I first characterise the variation that exists in the Soay sheep population at one cluster of MHC genes, known as class IIa, in a small number of sheep using genotyping methods which were developed for domestic sheep. Using this information, I was able to identify which alleles occur in close proximity on a single chromosome (a haplotype). I then use this information to develop a rapid method to

determine which pairs of haplotypes each individual carries for large numbers of individuals and applied it to 5349 individuals. Using this class IIa genotype data, I found evidence of balancing selection acting in this gene region to maintain alleles at frequencies that are more even than expected by chance. I also compared the MHC genotypes to five measures of individual health and productivity – body weight, three antibody measures (immunoglobulins A, E and G), and parasite burden. Many class IIa genotypes were related to variation in these measures of health and productivity, offering a glimpse into the evolutionary mechanisms which may operate to maintain high levels of diversity in this gene region.

# Declaration

This thesis is submitted in accordance with the requirements for a Doctorate of Philosophy by the School of Biological Sciences at the University of Edinburgh. The work included in this thesis has not been submitted for any other degree or professional qualification.

I declare that I am the sole author of this thesis. I conducted all of the analyses and wrote the thesis under the guidance of my supervisors. All work presented in this thesis is my own, with the following acknowledgements:

The long-term data and samples used throughout this thesis (including DNA samples analysed throughout, and body weight and parasite burden data analysed in Chapter 5) were collected as part of the Soay Sheep Project. I visited St. Kilda and assisted the project during lambing in spring 2014 and during the August catch in 2016.

DNA extractions for SNP analysis was carried out by Camillo Bérénos, Phil Ellis, and Liz Heap. Genetic parental assignments were carried out by Camillo Bérénos, and Jisca Huisman developed and implemented the Sequoia method to cluster half-siblings sharing ungenotyped parents. Susan Johnston merged the database and genetic pedigrees. The measure of inbreeding (Fhat3) used within chapter 5 was calculated by Camillo Bérénos.

Susan Johnston provided scripts to convert raw KASP genotypes into PLINK input format and to parse the BEAGLE output.

The immunoglobulin ELISAs which generated the data analysed in Chapter 5 were carried out by Kathryn Watt, Alex Sparks and Rona Sinclair, under the guidance of Dan Nussey and Tom McNeilly.

# Acknowledgements

This thesis would not have been possible without the hard work and support of many people to whom I am incredibly grateful. First and foremost, I would like to thank my incredibly supportive supervisors Josephine Pemberton and Keith Ballingall for your patience, support and encouragement throughout this PhD.

A huge thanks to the Soay sheep project, and particularly Jill, Josephine and Dan, for all the DNA samples, the immunoglobulin measures and the long-term data that made this PhD project possible. Many thanks to Jill Pilkington for teaching me how to catch lambs, to read the sheep and for the St. Kilda finishing school. To the 2013 and 2014 lambing teams and the 2016 August catch team for all the hill climbing and Puff Inn nights. To Charlotte Regan for enjoying the HUET training as much as I did, and then allowing me to join you as a field assistant and showing me the joys of watching frolicking lambs.

To Susan Johnston for the tremendous amount of support you've given me, all the incredibly well annotated R scripts and patience with my stats questions.

To lab 1.52 for all the troubleshooting help, especially Liz Heap. To Helen Todd at Moredun for helping me get my head around the cloning protocols and how not to ruin a polyacrylamide gel.

To the many members of WEG for teaching me pretty much everything I know about evolution and probably everything I never wanted to know about quantitative genetics.

To the Ashworth (& Roslin) sheepies, Charlotte, Al and Alex, for all the gin, wine, chocolate and of course talking through many sheep, parasite, immunological and stats questions.

To the amazing friends I've met along the way, particularly in office 1.33 and the 2013 cohort, for reminding me of the big wide world outside of Ashworth, for introducing me to the wonderful worlds of moths and sewing, for all the silly Friday afternoons that weren't always on a Friday, for all the many wonderful cakes, for so many enjoyable escapes to the pub and for the never-ending stream of WhatsApp messages.

To Lucy Webster for all your all your encouragement and support, to Kim Davie for all the long walks with the dogs to clear my head, and to you both for introducing me to so many different types of gin.

And finally, to my family for all your love and support throughout the years. It's been a long journey, but I promise I'll stop being a student now.

Thank you all.

# Table of Contents

# Abbreviations

| | |
|---|---|
| ABC | Associative balancing complex |
| AM | Animal model |
| cDNA | complementary DNA |
| DGGE | Denaturing gradient gel electrophoresis |
| FEC | Faecal egg counts |
| FS | Fluctuating selection |
| gDNA | genomic DNA |
| GIN | Gastrointestinal nematodes |
| GLMM | Generalised linear mixed effects model |
| HWE | Hardy-Weinberg equilibrium |
| Ig | Immunoglobulin |
| IPD | Immuno Polymorphism Database |
| LD | Linkage disequilibrium |
| MAF | Minor allele frequency |
| MHC | Major histocompatibility complex |
| NFDS | Negative frequency-dependent selection |
| PBR | Peptide binding region |
| PCR | Polymerase chain reaction |
| PMS | Pathogen-mediated selection |
| RFLP | Restriction fragment length polymorphism |
| RSCA | Reference strand-mediated conformation |
| SBG | Sequence-based-genotyping |
| SNP | Single nucleotide polymorphism |
| SSCP | Single strand conformation polymorphism |
| TCR | T cell receptor |

# 1. General introduction

Understanding how selective and neutral forces shape diversity within natural populations is a central aim of evolutionary population genetics. Most selective (e.g. positive and purifying selection) and neutral (e.g. drift) forces act to reduce variation in natural populations. Yet, natural populations typically exhibit high levels of variability. Haldane (1949) proposed the importance of pathogens in maintaining polymorphism, and some of the most variable genes are involved in immune responses to pathogens. The major histocompatibility complex (MHC) is one of the most variable regions in vertebrate genomes (Klein 1986; Parham & Ohta 1996; Garrigan & Hedrick 2003; Trowsdale 2011), which encode molecules involved in the detection of pathogens by the immune system. Pathogen-mediated balancing selection (PMS) is thought to be involved in the maintenance of these high levels of polymorphism at the MHC, and there is a growing body of evidence linking pathogens to MHC diversity (see Jeffery & Bangham 2000; Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). The exact evolutionary mechanisms through which balancing selection acts, however, are not well understood, and investigating these mechanisms requires extensive population data at both the phenotypic and genotypic levels.

Locus-specific genotype data is necessary to disentangle the various modes of balancing selection (Spurgin & Richardson 2010), but high levels of polymorphism and gene duplication events makes this challenging, especially in non-model organisms. MHC genes often occur in close proximity on a chromosome and are inherited as haplotypes, and so it may not always be possible to separate the effects of an allele at a particular locus from its haplotypic background. This thesis aims to address these challenges by 1) characterising the MHC class IIa haplotypic diversity in a well-characterised wild population, 2) develop and test a rapid haplotyping method, 3) haplotype large numbers of animals and 4) investigate associations between MHC haplotypes and phenotypes associated with health. Below, I summarise the structure and function of MHC genes, origin and maintenance of MHC diversity, the challenges in detecting and differentiating evolutionary mechanisms, introduce the study population and outline the main aims of this research.

## 1.1 MHC structure and function

The MHC is formed of clusters of genes with similar immune functions. MHC genes have been identified in all gnathostomes, or jawed vertebrates, in which they have been

looked for (Ohta *et al.* 2000; Kulski *et al.* 2002; Kelley, Walter & Trowsdale 2005). Up to three clusters of genes with similar functions are typically identified in the MHC, known as classes I, II and III (Klein 1986). In mammals, birds, reptiles and amphibians, the MHC classes are clustered together on a single chromosome, although not all clusters are always present and there is substantial variation in the number of genes within each cluster amongst taxa (Kelley *et al.* 2005). In teleost fish, however, the class I and class II loci are typically located on different chromosomes (Hansen *et al.* 1999; Sato *et al.* 2000; Kuroda *et al.* 2002).

The MHC encodes molecules that act as sentinels of infection, alerting the immune system to the presence of an invading pathogen and enabling it to mount an immune response. MHC class I and II molecules are membrane-bound glycoproteins that are capable of binding pathogen derived peptide fragments, and displaying them at the cell surface for recognition by appropriate T cells (Zinkernagel 1979; Murphy *et al.* 2012). Class I genes encode class I molecules which are expressed on all nucleated somatic cells and are primarily concerned with presenting endogenously derived peptides to CD8+ cytotoxic T cells. Recognition of the MHC peptide complex by the T cell receptor initiates a cytotoxic reaction resulting in the death of the target cell (Hughes & Yeager 1998). Class II genes encode class II molecules which are expressed on antigen presenting cells (B cells, macrophages and dendritic cells), and are primarily concerned with presenting exogenously derived peptides to CD4+ T cells. The CD4+ T cell subsets control and regulate both the antibody and the cellular responses to infection (Murphy *et al.* 2012). The class III subregion includes a large number genes, not all of which have immune related functions, and which are typically less variable than the genes in classes I or II (Kulski *et al.* 2002).

### 1.1.1 The *Ovar*-MHC

The MHC region in domestic sheep, known as *Ovar*-MHC for *Ovis aries*, is located on chromosome 20 q14 – q23 (Mahdy *et al.* 1989; Hediger, Ansari & Stranzinger 1991). Ruminants share a unique class II structure, which was first identified in cattle (Andersson *et al.* 1988; van Eijk *et al.* 1995), by which the class II is separated into class IIa and class IIb (Amills *et al.* 1998). The organisation of the three MHC classes is show in Figure 1.1. This organisation of the classes within sheep has been shown using linkage maps of microsatellites (Crawford *et al.* 1995; Maddox *et al.* 2001) and genomic sequencing of BAC clones (Liu *et al.* 2006, 2011; Qin *et al.* 2008; Gao *et al.* 2010).

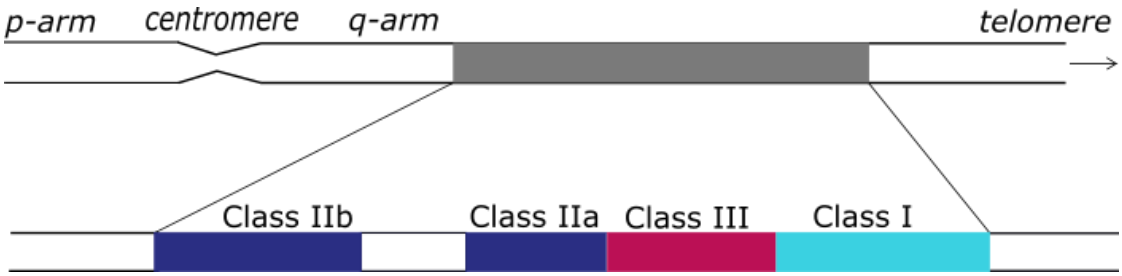



Figure 1.1. Schematic of the Ovar-MHC structure showing the relative position on chromosome 20. (Not drawn to scale)

### 1.1.2 Class I

The best characterised MHC region is the human leukocyte antigen (HLA) complex. Three highly polymorphic classical class I genes, A, B and C are present on all haplotypes along with three non-classical genes, E, F, G (Klein 1986). The majority of allelic diversity locates to exons 2 and 3 of the classical class I genes which form the peptide binding region (PBR) of the class I molecule. The class I molecule is a heterodimer consisting of three α domains and a $\beta_2$ microglobulin (Figure 1.2a), which is derived from a non-polymorphic locus external to the MHC (Hughes & Yeager 1998). The classical class I PBR binds peptides that are 9 amino acids long (Hughes & Yeager 1998). Non-classical class I genes are structurally similar to classical class I genes, but are relatively non-polymorphic and tend to show reduced or restricted expression patterns compared with the classical class I loci (Lawlor *et al.* 1990).

Class I molecules bind to peptides intracellularly and are transported to the cell surface where they are recognised by the antigen receptor of CD8+ cytotoxic T cells. Within the endoplasmic reticulum, class I molecules bind to cytosolic peptide fragments (Germain 1986), either a self-peptide (uninfected cells) or a non-self peptide (infected cells) which is derived from the breakdown of proteins from intracellular pathogens. After peptide loading, they are transported to the cell surface where they present the bound peptide. CD8+ T cells are continually circulating the body, and when they discover class I molecules loaded with a non-self peptide that is complementary to their T cell receptor (TCR), they form a complex which triggers a cytotoxic reaction to kill the infected cell. TCRs which bind to class I molecules presenting self-peptides are removed in the thymus by negative selection (Lawlor *et al.* 1990; Hughes & Yeager 1998) preventing autoimmune reactions. There may be some cross-presentation of

exogenously derived peptides by class I molecules (Kovacsovics-Bankowski & Rock 1995; Norbury *et al.* 1995; Yewdell, Norbury & Bennink 1999; Lizée *et al.* 2003; Ackerman & Cresswell 2004; Basha *et al.* 2012)

Orthologues of the HLA class I A, B and C loci cannot be identified in other species as class I genes appear to evolve rapidly, are typically derived by recent duplication and rearrangement events, and are highly polymorphic (Horton *et al.* 2004; Kelley *et al.* 2005). This makes it difficult to use information about class I genes in a model species to design assays for a non-model species. Cloning and sequencing has shown that there is a minimum of eight class I loci in Scottish blackface sheep, with variation in gene number between haplotypes (Miltiadou *et al.* 2005; Ballingall *et al.* 2008). Recent work to assemble the class I genome (Gao *et al.* 2010) and annotate it (Siva Subramaniam *et al.* 2015) have improved our knowledge of class I gene content and structure in sheep. Haplotype diversity in gene number within the class I region is not uncommon (Ellis *et al.* 1999; Karl *et al.* 2013), but these studies have been conducted on a very limited number of haplotypes.

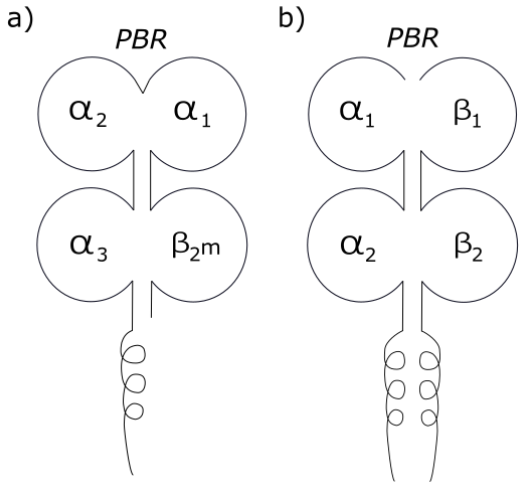



Figure 1.2. Schematic of structure of MHC class I molecules (a) and class II molecules (b), based on Hughes & Yeager (1998). $\beta_2$m refers to the $\beta_2$-microglobulin

### 1.1.3 Class II

Three classical HLA class II gene families, DP, DQ and DR are located within the human MHC (Klein 1986). Each contain closely linked pairs of A and B genes which encode the

α and β chains of the dimeric class II molecule (Figure 1.2b). The PBR is formed between the $\alpha_1$ and $\beta_1$ domains (Figure 1.2b), and can bind peptides of a more variable size range, typically between 11 and 17 amino acids long (Hughes & Yeager 1998). Class II molecules are primarily expressed on antigen presenting cells – dendritic cells, macrophages, B cells and thymic epithelial cells (Hughes & Yeager 1998; Ting & Trowsdale 2002; Jensen 2007).

Like class I molecules, the class II molecules are loaded with a peptide intracellularly (see Figure 1.3a), but with the assistance of a non-classical class II molecule, DM (Kropshofer, Hämmerling & Vogt 1997). The peptide MHC complex is transported to the cell surface for recognition by CD4+ T helper cells. Shortly after synthesis, class II molecules are bound to an invariant chain, which prevents binding to intracellular peptides, and chaperones the class II molecule to MIIC lysosomal-like vesicles (Ting & Trowsdale 2002). The invariant chain degrades, and the class II molecule can bind to degraded exogenously-derived peptides which have been internalised via phagocytosis or receptor-mediated endocytosis (Jensen 2007). Once bound, the class II–peptide complex is transported to the cell surface where presentation to circulating antigen specific CD4+ T cells takes place. Following T cell activation an adaptive immune response is initiated (see Figure 1.3b) through the release of cytokines and chemokines which can stimulate inflammatory and antibody type responses to control the infection (Murphy *et al.* 2012).

Of the three MHC classes, class II is the best characterised in sheep (Amills *et al.* 1998; Dukkipati *et al.* 2006a). Class II genes typically show a high degree of homology amongst species, and orthologues of the DQ and DR genes can readily be identified in mammals (Garrigan & Hedrick 2003; Kelley *et al.* 2005). This has facilitated the study of class II genes in a range of species including ruminants. Gene nomenclature for class II follows that used in the human MHC, where the first two letters denote the locus (*DR* or *DQ*), followed by whether it is the A or B gene (e.g. DRA and DRB), and finally the gene number if the locus is duplicated (e.g. DRB1).

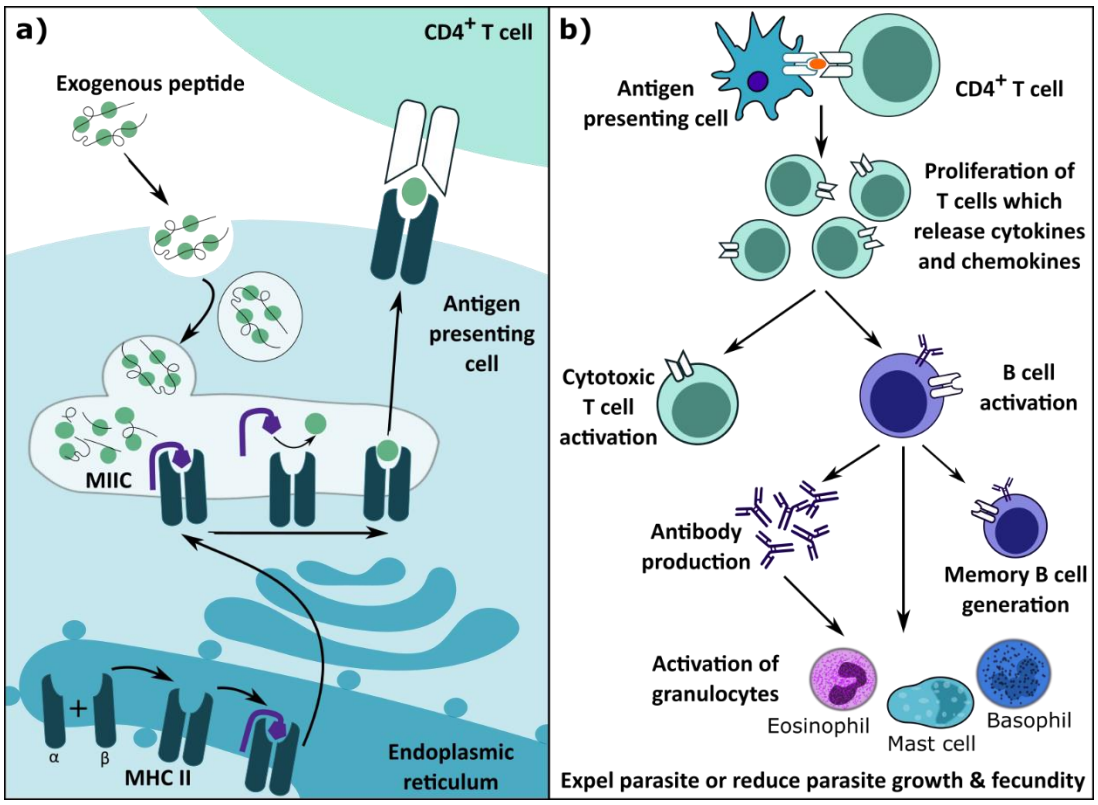



Figure 1.3. a) MHC class II molecules are produced and loaded with exogenously derived peptides within antigen processing cells. The α and β chains are produced within the endoplasmic reticulum where the peptide binding site becomes associated with an invariant chain protein, before moving into a specialised vesicle (MIIC). Exogenously derived proteins are engulfed and transported to the MIIC vesicle, where they are broken down. Within the MIIC, the invariant chain is cleaved from the MHC class II molecule and replaced with an exogenously derived peptide fragment. This MHC II – peptide complex is transported to the cell surface for recognition by the antigen specific receptors where it forms a complex with a T cell receptor on $CD4^+$ T cells. The cascade of events following T cell recognition of the MHC-peptide complex is summarised in (b). The activation and proliferation of T cells supports immune responses through the production of a range of pro- and anti-inflammatory cytokines and chemokines which control and regulate immune effector cell populations including cytotoxic T cells, B cells, which generate antibodies, memory B cells and recruit granulocytes (eosinophils, basophils and mast cells) which expel parasites or reduce their growth and fecundity. Figure 1.3a adapted from Neefjes *et al.* (2011). Figure 1.3b created by author from information in Murphy *et al.* (2012).

Class IIa in sheep includes the classical class II genes which encode the *DR, DQ1* and *DQ2* molecules (Figure 1.4), as well as a number of pseudogenes and non-classical genes (Dukkipati *et al.* 2006a). There is a single transcribed *DRA* locus with low polymorphism (Ballingall *et al.* 2010; Subramaniam *et al.* 2012), which combines with a single expressed and highly polymorphic *DRB1* locus (Deverson *et al.* 1991; Scott *et al.* 1991; Ballingall *et al.* 1992). Both the *DQA* and *DQB* loci are duplicated in sheep

(Chardon *et al.* 1985; Scott, Choi & Brandon 1987; Scott *et al.* 1991; Fabb *et al.* 1993; van Oorschot *et al.* 1994; Wright & Ballingall 1994; Ballingall *et al.* 2015, 2017 in press). Most haplotypes carry the *DQ1/DQ2* configuration (as shown in Figure 1.4); however, the *DQA1* locus cannot be identified on some haplotypes (Fabb *et al.* 1993; Ballingall *et al.* 2015; Ali *et al.* 2016), termed *DQA1* null. Similarly, the *DQB1* cannot be identified on all haplotypes (Ballingall *et al.* 2017 in press), termed *DQB1* null. Haplotypes lacking the *DQA1* and *DQB1* loci were typically found in combination with two *DQA2* and two *DQB2* loci, the additional loci termed *DQA2-like* and *DQB2-like* (Hickford, Ridgway & Escayg 2000; Hickford *et al.* 2004; Hickford, Zhou & Fang 2007; Ballingall *et al.* 2015, 2017 in press). It is unknown whether the *DQA2-like* and *DQB2-like* alleles are truly located at a third new locus, or are highly divergent alleles of the *DQA1* and *DQB1*, although Ali *et al.* (2016) identified haplotypes carrying all three alleles for each of the *DQA* and *DQB* loci. Ballingall (in press) recently showed that *DQA2-like* and *DQB2-like* transcripts were only able to co-express with one another, suggesting that only haplotypes carrying both *DQ2-like* genes will be able to generate a functional class II molecule. No equivalent of the HLA-DP genes has been identified in sheep or other ruminants.

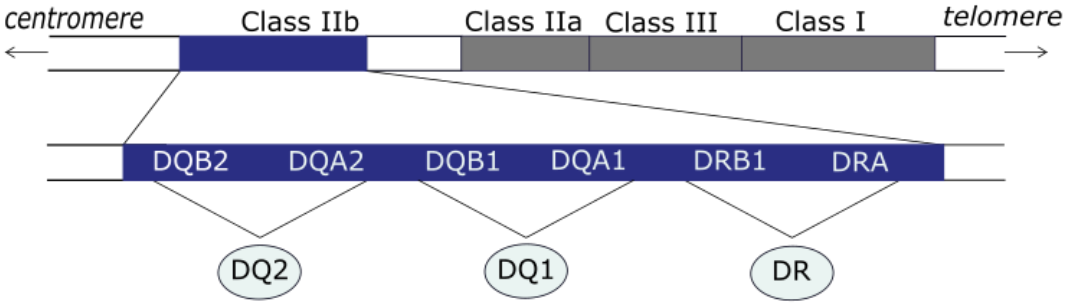



Figure 1.4. Class IIa genes as they are organised on sheep chromosome 20. The molecules formed by each pair of loci are shown below. Note that only the typical DQ haplotype structure is shown. (Not drawn to scale)

Class IIb genes encode the ruminant specific *DY* molecules, as well as the non-classical *DN/DO* and *DM* molecules (Amills *et al.* 1998; Dukkipati *et al.* 2006a), all of which have low polymorphism (Dukkipati *et al.* 2006a). The *DYA* and *DQB* genes have been identified in sheep (Wright, Ballingall & Redmond 1994; Ballingall & McKeever 2005), cattle (Andersson *et al.* 1988; Stone & Muggli-Cockett 1990; van der Poel *et al.* 1990;

Ballingall *et al.* 2001, 2004) and goats (Mann *et al.* 1993) and appear to be unique to ruminants. *DYA* and *DYB* loci show limited expression in ruminants and are probably restricted to dendritic cells (Ballingall *et al.* 2001; Ballingall & McKeever 2005). *DNA* and *DOB* probably form a single pair of genes, but their expression in sheep is uncertain (Wright, Redmond & Ballingall 1995). The human homologue, *DO*, is expressed in B cells and is thought to form complexes with *DM* molecules to enhance the ability of B cells to present antigens internalised via the B cell antigen receptor (Jensen 2007). *DMA* and *DMB* genes have been identified in sheep (Gao *et al.* 2010; Lee *et al.* 2012), although their expression in sheep does not appear to have been assessed. The human homologue of *DM* is expressed in antigen presenting cells and does not bind peptides. Instead, it interacts with classical class II molecules during peptide loading, and it is thought that the DM molecules are involved in ensuring that stable class II-peptide complexes are presented on cell surfaces by exchanging peptides and favouring the formation of more-stable complexes (although there is some uncertainty surrounding this function) (Jensen 2007).

## 1.2 Origins and diversity at the MHC

### 1.2.1 Evolutionary origins of the MHC

An MHC region, a complex of linked genes involved in an adaptive immune response, is only found in jawed vertebrates (Kelley *et al.* 2005). Within jawed vertebrates, the number and content of genes that comprise the MHC varies across and within taxa (Kelley *et al.* 2005), but typically include those with classical peptide-binding functions, housekeeping genes and MHC assembly genes. Many MHC genes with non-peptide-binding functions have orthologues which can be found in plants, invertebrates and vertebrates as they relate to general cell organelle functions (Kelley *et al.* 2005), and such genes have been co-opted from existing genes with similar function, such as the TAP1 and TAP2 genes which originated in the adenosine triphosphate binding cassette family of genes (Danchin *et al.* 2004). Genes with classical MHC peptide presentation functions likely evolved after divergence of jawed and jawless vertebrates, though homologues of these genes have not been identified in non-jawed vertebrates, making their origins obscure (Klein & O'hUigin 1993; Kelley *et al.* 2005; Flajnik & Kasahara 2010).

The organisation of the MHC into three regions with similar functions (classes I, II and III as discussed above) is typical of mammals, and the presence of at least two clusters has been identified in many jawed vertebrates (Kelley *et al.* 2005). These clusters likely originated from a duplication event, probably two whole-genome duplication events, prior to the radiation of jawed vertebrates (Flajnik & Kasahara 2010). The organisation of these clusters and their gene content differs across taxa (Kelley *et al.* 2005). For example, chickens have a "minimal essential" MHC with only 19 genes (Kaufman *et al.* 1999), whereas songbirds may have hundreds of genes (Bollmer *et al.* 2010; Sepil *et al.* 2012), and in teleost fish, the class I and II loci do not always seem to be linked (Shum *et al.* 2001).

### 1.2.2 Generation of new MHC alleles

Allelic diversity at an MHC locus can be generated by point mutations: substitutions, insertions or deletions. These mutations will either have no effect (e.g. synonymous mutations) or alternatively be advantageous or disadvantageous. Either way, they accumulate to generate the high levels of allelic diversity observed within the MHC. There is no evidence, however, that the mutation rate is higher at the MHC than elsewhere in the genome (Hughes & Hughes 1995).

The MHC is also characterised by high gene diversity. The region typically has high gene density, high similarity between genes (i.e. alleles at different loci are similar) and the retention of allelic lineages over long periods of evolutionary time (Figueroa, Günther & Klein 1988b), such that alleles from one species may be more similar to alleles of another species than within the same species, known as trans-species polymorphism (TSP) and has been identified in many taxa (Miller & Withler 1997; Hoelzel, Stephens & O 'brien 1999; Garrigan & Hedrick 2001; Van Den Bussche, Ross & Hoofer 2002; Richardson & Westerdahl 2003; Musolf, Meyer-Lucht & Sommer 2004; Alcaide, Edwards & Negro 2007; Eimes *et al.* 2015), including ruminants (Ballingall *et al.* 2010). Two primary mechanisms have been proposed to explain the generation of gene diversity at the MHC, concerted evolution (Lawlor *et al.* 1990; Ohta 1999) and birth-and-death evolution via gene duplication (Nei & Hughes 1992; Ota & Nei 1994), only the latter of which is compatible with trans-species polymorphism. Birth-and-death evolution and gene conversion may not be mutually exclusive events (Edwards *et al.* 1999; Wittzell *et al.* 1999), but rather occur on different time scales.

Concerted evolution within the MHC is the creation of new alleles through gene conversion (non-reciprocal exchange of short DNA segments) or inter-locus recombination (Lawlor *et al.* 1990; Ohta 1999), and the resulting genes evolve together so that genes within a species are more similar to one another than between species (Hess & Edwards 2002). Gene conversion at MHC loci has been identified in humans (Watkins *et al.* 1992; Belich *et al.* 1992; McAdam *et al.* 1994; Zangenberg *et al.* 1995; Marcos *et al.* 1997; Högstrand & Böhme 1999), other mammals (Andersson *et al.* 1991; She *et al.* 1991; Schaschl *et al.* 2005), birds (Hunt, Pharr & Bacon 1994; Wittzell *et al.* 1999; Miller & Lambert 2004a; Chaves *et al.* 2010; Spurgin *et al.* 2011; Gillingham *et al.* 2016) and fish (Langefors, Lohm & von Schantz 2001b; Reusch & Langefors 2005). Gene conversion can lead to reduced locus number and allelic diversity (Gu & Nei 1999).

Birth-and-death evolution (Nei & Hughes 1992; Ota & Nei 1994) is a processes whereby gene duplications create new genes, and while some persist and evolve, others are deleted or degrade and become pseudogenes (Nei & Rooney 2005). Nei *et al.* (1997) and Gu & Nei (1999) showed that alleles at an MHC locus typically cluster together phylogenetically, and this has been shown many times since (Go *et al.* 2003; Piontkivska & Nei 2003; Axtner & Sommer 2007). Birth and death evolution of MHC alleles seems to be more common in mammals than birds (Hess & Edwards 2002), and though it may not be a ubiquitous process in birds (Edwards *et al.* 2000; Hess *et al.* 2000), there is good evidence it does occur (Bollmer *et al.* 2010). Class II gene families tend to be similar across mammalian species, suggesting that duplication events are typically very ancient, and the loci involved have subsequently undergone divergent evolution. A probable recent gene duplication event of *DRB* genes was identified in the bank vole, *Clethrionomys glareolus* (Axtner & Sommer 2007). The presumed duplicated *DRB* genes have low polymorphism compared to the presumed original genes (Axtner & Sommer 2007) and have not been identified in another arvicolid species, the water vole, *Arviola terrestris* (Oliver & Piertney 2006).

## 1.3 Maintenance of diversity through balancing selection

Understanding functional variation is an important field of research in evolutionary biology. Polymorphisms at adaptive loci rarely persist within populations as, typically, the most advantageous allele increases to fixation. Balancing selection is the

maintenance of multiple alleles at a locus and can maintain both low and high numbers of alleles. The prevalence of balancing selection is relatively unknown, and may be restricted to a relatively small number of loci (Fijarczyk & Babik 2015). A classic example of balancing selection maintaining low levels of polymorphism is the persistence of the wild-type and sickle-cell alleles at the *β-globin* locus in humans, which causes sickle-cell anaemia in homozygous form but is protective against malaria in its heterozygous form (Allison 1956). Another example is the persistence of horn size polymorphism in Soay sheep (*Ovis aries*) at the bi-allelic locus *RXFP2*, where the allele for larger horns confers greater reproductive success and the allele for smaller horns confers increased survival, resulting in the heterozygote possessing greatest fitness (Johnston *et al.* 2013). Both sickle-cell haemoglobin and Soay sheep horn type are maintained due to the greater fitness of the heterozygotes, a specific form of balancing selection known as heterozygote advantage (Hedrick 2012). In plants, high levels of polymorphism are found at self-incompatibility (*S*) loci, a complex of linked genes or gene groups, which prevent self-fertilisation in Angiosperms (De Nettancourt 2001). These loci encode recognition proteins that recognise and reject pollen from genetically similar individuals, thus promoting outcrossing (Silva & Goring 2001). Thus, multiple alleles must be maintained within the population to facilitate cross-fertilisation, and indeed, *S*-loci are highly polymorphic (Charlesworth & Awadalla 1998; Charlesworth *et al.* 2000). *S*-loci diversity is maintained by negative-frequency dependent selection, whereby rare alleles have an advantage as they are less likely rejected as "self" alleles (Charlesworth *et al.* 2000; Llaurens *et al.* 2008; Schierup & Vekemans 2008). Within vertebrates, and particularly within mammals, the MHC contains highly polymorphic loci, the diversity of which is maintained by balancing selection, acting either through natural or sexual selection. PMS is often invoked as the primary form of natural selection likely to be acting on the MHC (Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). Sexual selection may also act through disassortative mating to optimise or maximise offspring genetic diversity (Milinski 2006; Piertney & Oliver 2006).

### 1.3.1 Natural selection

Functional MHC polymorphism is greatest in the PBR coding regions (Hughes & Hughes 1995; Sommer 2005). The PBR structure determines which peptides an MHC isoform can detect and present to the immune system (Murphy *et al.* 2012). Thus, the array of MHC isoforms an individual possesses determines which pathogen derived

peptides (both at the species and genotypic level) can be presented to the immune system and detected. The ability to detect and respond to pathogen infection is likely to impact individual fitness, for example through survival or reproductive success. Natural selection should therefore act on MHC alleles, so as to promote variation through balancing selection, the selective maintenance of multiple alleles at a locus within a population (Hughes & Yeager 1998; Meyer & Thomson 2001; Hedrick 2002; Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). There are four modes of natural selection that have been proposed, the first three of which pertain to PMS (heterozygote advantage, negative frequency-dependent selection and fluctuating selection) and the fourth to deleterious mutations (associative balancing complex - ABC). These mechanisms may not be mutually exclusive (Spurgin & Richardson 2010).

Under heterozygote advantage (Doherty & Zinkernagel 1975) selection favours individuals with the ability to recognise and react to the broadest range of peptides; therefore individuals with the greatest variety of alleles, heterozygotes, should be fittest. Only overdominant heterozygote advantage, where the heterozygote is fitter than both homozygotes, is capable of maintaining MHC diversity (Takahata & Nei 1990). Dominant heterozygote advantage in which the heterozygote is fitter than the mean of the two homozygotes, is not capable of contributing to the maintenance of MHC diversity. Heterozygote advantage is not expected to operate in single infection scenarios, which are typical of much experimental work (Apanius *et al.* 1997), but is more likely in co-infection or serial infection situations (Doherty & Zinkernagel 1975; Hughes & Nei 1992) which are more typical of most wild systems. Many models of heterozygote advantage assume symmetry in selection across all homozygotes and heterozygotes, which may be inherently unlikely at first sight (De Boer *et al.* 2004; Stoffels & Spencer 2008). Theoretical modelling has shown that symmetrical selection could generate observed levels of MHC variation (Maruyama & Nei 1981; Takahata & Nei 1990), whilst asymmetric selection could not (Lewontin, Ginzburgh & Tuljapurkar 1978; Hedrick 2002; Borghans, Noest & De Boer 2003; De Boer *et al.* 2004). Pathogens contain many epitopes, each of which could be recognised by different MHC molecules, so there may be much overlap in the pathogen sets that can be recognised by different molecules (see Stoffels & Spencer 2008 and references within). Stoffels & Spencer (2008) have shown that when this degeneracy within MHC molecules is considered, a

symmetrical heterozygote advantage may not be unrealistic and could maintain diversity.

Heterozygote advantage might be expected to favour even greater rates of gene duplication and diversification of gene families rather than just increasing, or maintaining, allelic diversity (Potts & Wakeland 1990). In mammals, there is arguably a pattern for increased gene diversity within class I, but less so for class II which shows relatively restricted classical gene diversity (Kelley *et al.* 2005). Nowak *et al.* (1992) suggested that high diversity would lead to self-reactivity or redundancy as the matching T cell receptors would be negatively selected in the thymus. Thus, there may be an optimum MHC diversity (Nowak *et al.* 1992; Woelfing *et al.* 2009), and studies from non-mammalian species that exhibit variation in gene number have often found that an intermediate number of alleles or intermediate allelic diversity is most advantageous (e.g. fish - Wegner *et al.* 2003b; Forsberg *et al.* 2007; Wegner 2008; Kalbe *et al.* 2009; McCairns *et al.* 2011; Hablützel *et al.* 2014; birds - Hawley & Fleischer 2012; reptiles - Madsen & Ujvari 2006; see Appendix A). Whilst this does not resolve whether or not heterozygote advantage operates within such populations, it does suggest that there is some limit to the optimum number or diversity of genes and alleles.

Under negative frequency-dependent selection (NFDS), also known as rare-allele advantage (Bodmer 1972; Takahata & Nei 1990), a cyclical co-evolutionary arms race between parasites and their hosts is envisioned. Parasites possessing antigens unrecognised by the most common host genotypes will be able to spread through the host population. Host individuals carrying new or rare alleles that allow identification of those parasites will have a selective advantage, but only until the allele becomes common (Takahata & Nei 1990). The allele then loses its selective advantage, and the parasite strain becomes less prevalent. The cycle begins again with a new parasite strain and corresponding MHC allele. New alleles (Takahata & Nei 1990) or rare old alleles (Slade & McCallum 1992) could both be selectively advantageous, although it may take up to 1000 generations for an allele that has lost its selective advantage to become rare again (Apanius *et al.* 1997). Monitoring allele frequencies in a population over time is the only way to truly test NFDS, yet monitoring populations over evolutionary time is unfeasible (Piertney & Oliver 2006). Additionally, rare alleles are more likely to occur in heterozygotes (Spurgin & Richardson 2010).

Under fluctuating selection (FS) (Hedrick, Thomson & Klitz 1987; Hedrick 2002) alleles vary in their selective advantage according to diversity within the parasite populations.

Parasite populations varying in space and time due to the environment external to the host will exert differing selection pressures on host MHC alleles across space and time (Spurgin & Richardson 2010). This mechanism is notably different from NFDS in that the allele frequency changes are driven by extrinsic factors, for instance changes in the pathogen populations which are unrelated to MHC diversity (Meyer & Thomson 2001). Hedrick and Kim (2000) showed that theoretically, fluctuating selection could generate observed frequencies of alleles even after accounting for temporal autocorrelation.

A fourth model of MHC evolution, ABC, proposed by van Oosterhout (2009) suggests that, if a certain degree of diversity already existed, deleterious mutations within the regions surrounding MHC genes contribute to the maintenance of polymorphisms. This model proposes that the mutations are not purged by selection because they occur within MHC linkage blocks, within which there is little recombination resulting in high linkage disequilibrium (LD), and are rarely expressed in homozygous form because of the diversity of MHC alleles. Thus, the deleterious impact of such mutations is reduced and purifying selection is inefficient (Charlesworth & Willis 2009). These mutations therefore become fixed and contribute to balancing selection. In a recent study on the human MHC, Lenz *et al.* (2016) have shown that there is an increase in the numbers of putatively deleterious variants within the MHC region, which decline in frequency with distance from the MHC.

### 1.3.2 Sexual selection

MHC-based mate choice is an alternative mechanism widely suggested as promoting polymorphism at the MHC (Potts & Wakeland 1990; Brown & Eklund 1994; Jordan & Bruford 1998; Zelano & Edwards 2002; Milinski 2006). The ability to choose a mate based on the MHC alleles they possess could enable an individual to maximise its offspring's fitness under two, non-exclusive strategies: maximising immunocompetence or avoiding inbreeding. Choosing an MHC dissimilar or MHC diverse mate would maximise MHC allelic diversity in offspring and thus their ability to detect a range of parasites (Penn & Potts 1999); although a complementary number of alleles to an individual's own in order to reach an optimum diversity may be preferable (Aeschlimann *et al.* 2003). Inbreeding reduces fitness (Reed & Frankham 2003), thus avoiding mating with kin by choosing an MHC dissimilar mate could maximise offspring fitness both at the MHC and across the genome (Penn 2002).

Early studies on MHC-based mate choice in mice (reviewed in Jordan & Bruford 1998) and humans (Ober *et al.* 1997; Génin *et al.* 2000) found evidence for disassortative mating, usually via female preference (Meyer & Thomson 2001), although not all studies were able to identify disassortative mating (Hedrick & Black 1997). With improved statistical and genotyping methods, studies have revisited this question in a range of species, some finding evidence supporting its occurrence (Reusch *et al.* 2001; Radwan, Tkacz & Kloch 2008; Miller *et al.* 2009; Huchard *et al.* 2010; Juola & Dearborn 2011; O'Farrell *et al.* 2012; Reichard *et al.* 2012; Strandh *et al.* 2012; Løvlie *et al.* 2013; Winternitz *et al.* 2013; Yang *et al.* 2014) and others, not (Knafler *et al.* 2012; Niskanen *et al.* 2016).

MHC-based mate choice relies on an individual's ability to detect a mate's MHC alleles and possibly their own, and recent evidence has shown that this may be possible in a number of species (Ziegler *et al.* 2010; Milinski *et al.* 2013). MHC disassortative mating does not, however, easily explain the extensive evidence for diversifying selection found specifically within the PBR, where diversity is related to presentation of peptides, rather than across the entirety of the locus (Hughes & Hughes 1995) (see next section). That is not to say that MHC disassortative mating could not be operating in conjunction with other modes of selection, but these functional patterns of diversity associated with peptide presentation seem more likely to be due to a mechanism related to MHC function, such as PMS.

## 1.4 Detecting selection at the MHC

### 1.4.1 Historical selection

If positive and/or balancing selection have acted on a species or population in the past, there should be detectable signatures within the nucleotide sequences. The most commonly applied tests of historical selection in the case of the MHC are dN/dS, trans-species polymorphism, the Ewens-Watterson tests, and patterns of differentiation amongst spatially separated populations, although others are discussed in more detail elsewhere (Bernatchez & Landry 2003; Garrigan & Hedrick 2003; Piertney & Oliver 2006). Whilst the Ewens-Watterson test assesses evidence of recent selection, the dN/dS and trans-species polymorphism tests assess signatures of selection that can last for very long periods of time. Garrigan and Hedrick (2003) demonstrated that whilst these signals of selection can be generated over relatively few generations, they may

take tens or even thousands of generations to degrade. Such tests should therefore be interpreted as evidence of historical balancing selection acting upon a species, that may or may not continue within contemporary populations.

Selection can be detected by assessing the patterns of synonymous and nonsynonymous mutations (dN/dS) within exonic nucleotide sequences (Hill & Hastie 1987; Hughes & Nei 1988). Synonymous mutations are, for the most part, effectively neutral, whereas non-synonymous mutations can alter the conformation and binding capabilities of the MHC protein. Non-synonymous mutations are therefore assumed to be present within the sequence because of selection. The rate of synonymous mutations is assumed to be approximately equal to the rate of mutation. If the rate of non-synonymous mutations (dN) exceeds synonymous mutations (dS), $dN/dS > 1$, then it can be assumed that positive selection has acted (Bernatchez & Landry 2003). Such calculations are typically carried out using maximum likelihood estimates first developed by Felsenstein (1981) which incorporate mutational models, such as invariable, neutral and positively selected sites (Nielsen & Yang 1998; Yang et al. 2000; Suzuki 2004; Massingham & Goldman 2005). This test of selection, unlike many others, makes no assumptions about population size, population structure or whether the population is at equilibrium (Nielsen 2001). Most studies that test for dN/dS identify a signal of positive selection within the PBR (Bernatchez & Landry 2003; Piertney & Oliver 2006). The ubiquity of this result suggests that those studies that do not identify positive selection may not have correctly identified the PBR.

Trans-species polymorphism is the maintenance of alleles across speciation events (Figueroa, Gúnther & Klein 1988a), as discussed above, and is an indicator or balancing selection. Under neutral and coalescent theories, the number of alleles shared between closely related species would be expected to decline with time following speciation (Piertney & Oliver 2006), but balancing selection can retain alleles for much longer (Takahata & Nei 1990). Phylogenetic comparison of MHC alleles amongst related species and detection of incomplete lineage sorting of MHC alleles at a locus is a signature of balancing selection (Bernatchez & Landry 2003; Piertney & Oliver 2006). The absence of trans-species polymorphism, though, may not be indicative of the absence of balancing selection if concerted evolution has occurred (Piertney & Oliver 2006).

The Ewens-Watterson test (Ewens 1972; Watterson 1978) compares the observed allele frequencies to those expected under neutrality to infer which process is most

likely operating. The test makes the assumption that the test population has remained the same size historically and at equilibrium. To avoid confounding by such demographic effects, patterns observed for MHC alleles are frequently compared to those of neutral markers (Garrigan & Hedrick 2003). Solberg *et al.* (2008) performed a meta-analysis of 497 human populations and found that the Ewens-Watterson test identified balancing selection at all but one HLA locus.

Patterns of population differentiation or similarity at MHC markers which are not reflected by neutral markers can also signify selection (Bernatchez & Landry 2003). Greater differentiation at the MHC compared to neutral markers suggests that selection is acting differently in different populations due to variation in selection pressures (Jeffery & Bangham 2000). For example, if pathogen pressures varied between populations, the selective advantages of an MHC allele may vary between populations. Reduced differentiation at MHC compared to neutral markers indicates that, despite demographic processes acting to generate differences between populations, balancing selection is maintaining similar MHC differentiation (Schierup, Vekemans & Charlesworth 2000). Muirhead (2001) showed theoretically that population differentiation (measured as $F_{ST}$) at MHC loci is reduced amongst subdivided populations under balancing selection compared with neutral markers for varying levels of selection and migration. Neutral processes acting within current or recent generations may outweigh any effects of selection, particularly large demographic effects such as a bottleneck. Numerous studies have investigated patterns of MHC differentiation within or between species in natural systems (see Appendix A and reviews – Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010), but it can be very difficult to disentangle the different processes of selection, demography and social structure acting in each of the populations to truly exclude demography or social structure as explanations for the observed patterns (Piertney & Oliver 2006).

### 1.4.2 Contemporary selection

An excess of heterozygotes would be expected under balancing selection if either heterozygote advantage (Doherty & Zinkernagel 1975) or NFDS is operating (Penn 2002). Under both these forms of balancing selection, heterozygotes would be expected to be at a selective advantage either due to the heterozygosity itself, or the presence of advantageous rare alleles. Some studies in humans have identified heterozygote excess

(Black & Salzano 1981; Markow et al. 1993; Chen et al. 1999) but few studies in other species have been able to detect this (Spurgin & Richardson 2010). An excess of heterozygotes was observed in a neotropical rodent, Talas tuco-tuco *Ctenemys talarum*, and heterozygote advantage was excluded through a lack of association between heterozygosity and parasite load (Cutrera, Zenuto & Lacey 2011). Often, however, selection may be too weak to be detected within a single generation (Garrigan & Hedrick 2003; Spurgin & Richardson 2010).

Selection against homozygotes or particular genotypes may manifest before birth or hatching, or in very early life which would result in allele frequencies deviating from Mendelian segregation proportions (Garrigan & Hedrick 2003). However, testing for deviations from Mendelian proportions is challenging as it requires very large sample sizes, MHC genotypes for both parents and all observed offspring. These requirements may be met within human populations, as discussed by Garrigan & Hedrick (2003) and observed by Black & Hedrick (1997) in South American Amerind families. However, the necessary data is typically very difficult to obtain from wild populations, where mating attempts are difficult to observe, where females may be promiscuous, and where unobservable post-copulatory mechanisms determine paternity. Evidence for deviations from Mendelian proportions were identified in lesser kestrels, *Falco naumanni* (Alcaide *et al.* 2012) and domestic dogs, *Canis familiaris* (Niskanen *et al.* 2016), but not in an inbred colony of rhesus macaques (Sterck *et al.* 2017).

Identification of associations between MHC alleles and disease or fitness traits is the only direct way to measure natural selection acting in contemporary populations (Potts & Wakeland 1990; Apanius *et al.* 1997). Natural selection on MHC alleles acts through the phenotypes they confer, such as resistance or susceptibility to disease or their impact on individual fitness. Associating pathogen prevalence (qualitative resistance) or burden (quantitative resistance) with MHC variation demonstrates contemporary PMS, and, depending upon the type of data collected, enable modes of balancing selection to be tested. For example, heterozygote advantage would be supported by associations between pathogen diversity and heterozygosity, and rare-allele advantage or fluctuating selection by associations between specific alleles and specific pathogens (Spurgin & Richardson 2010). Such associations may, however, be difficult to detect when selection is weak, selective pressures vary amongst co-infecting pathogens or across spatio-temporal scales, or when allelic diversity is high and sample sizes are small (Apanius *et al.* 1997). Small differences in the ability to resist disease may be

amplified in lifetime measures of fitness (Potts & Wakeland 1990), making it more feasible to detect associations (Apanius *et al.* 1997). Many studies have found associations between MHC variation and pathogen or fitness measures and have been reviewed extensively (see Apanius *et al.* 1997; Jeffery & Bangham 2000; Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010; Trowsdale 2011; Appendix A).

## 1.5 Challenges in detecting and differentiating modes of selection

Despite hundreds of MHC association studies in a wide range of species (see Appendix A), few studies have been able to attribute observed patterns of variation to any one of the mechanisms of balancing selection. Spurgin and Richardson (2010) discussed the challenges in doing so, and showed that the predictions we can make to explain observations obtained from many of the detection methods are almost always compatible with multiple PMS explanations. An excess of heterozygosity, for example, could be compatible with both heterozygote advantage and NFDS (Penn 2002). Associations between specific MHC alleles and disease may be compatible with NFDS and FS (Spurgin & Richardson 2010). Reduced differentiation between populations at MHC compared to neutral loci may be compatible with NFDS or heterozygote advantage, and greater differentiation between populations at MHC compared to neutral loci could be attributed to NFDS or FS (Spurgin & Richardson 2010). Many studies still infer one mechanism over another without being able to exclude the alternative (Spurgin & Richardson 2010). It is possible, and perhaps likely, that multiple mechanisms operate within a single population, and that a number of sampling, technical and statistical limitations have frequently hampered the ability of studies to disentangle the different mechanisms (Spurgin & Richardson 2010).

Pathogen sampling strategies must also sufficiently represent the range of pathogen pressures acting on the population and should ideally encompass adequate temporal and spatial scales (Spurgin & Richardson 2010). Sampling a single pathogen may reveal the selection imposed by that one pathogen, but wild populations are typically co-infected or serially infected by multiple pathogens (Schmid-Hempel 2011). Intensive sampling of parasite communities may therefore be valuable for MHC association studies, for example a recent study used next generation sequencing of host bacterial

communities as a measure of host pathogen diversity (Collin *et al.* 2013). Not all pathogens exert strong selection pressure, so ideally sufficient groundwork should also be carried out to understand the associations between pathogens or pathogen communities and fitness. Small differences in pathogen resistance may be amplified in individual fitness (Potts & Wakeland 1990), and therefore fitness may be more revealing of the variation in selection pressures acting on MHC variation (Apanius *et al.* 1997). Testing multiple phenotypic traits, including but not restricted to qualitative and quantitative measures of pathogen infections, may also help to generate an improved picture of selection. Furthermore, studies often take a "snapshot" of the allele frequencies within a population at a single location, thus it is not possible to assess changes in genotype frequencies over time or space which would likely occur under NFDS or FS (Spurgin & Richardson 2010). The long-term intensive monitoring of spatially distributed wild populations required for such an assessment, may not realistically be feasible, however.

Applying adequate genotyping methods which attribute alleles to loci and reflect functional variation are a key technical challenge when studying the MHC in non-model species (Bernatchez & Landry 2003; Spurgin & Richardson 2010). Characterisation of each locus individually, rather than co-amplification of multiple indistinguishable loci, is required to partition variation amongst loci and determine an individual's true genotype. With only allele counts for an individual, it is not possible to apply most analysis methods to discriminate modes of selection, for example heterozygote advantage could not be detected without knowing an individual's heterozygosity (Spurgin & Richardson 2010). To understand modes of selection it is important to consider expressed loci, or allelic variation may be overestimated and will not reflect only functional variation. Genotypes at the locus itself should be considered, rather than those of flanking loci such as microsatellites. Whilst a flanking microsatellite might be in LD with the locus of interest, lack of selective constraint and highly repetitive sequences enables variation to build up more rapidly (Ellegren 2004) so that multiple microsatellite alleles may be in LD with a single functional allele (Paterson 1998).

Finally, a multi-locus system such as the MHC should not be assessed using a single locus. Many alleles at multiple loci is statistically difficult to deal with because each genotype has low sample size and large numbers of alleles can result in a lack of statistical power (Hill 1991). If loci are analysed separately, this may lead to an increased chance of type II error. How to apply corrections for multiple testing then

becomes a key question. Additionally, these loci are non-independent as there is high LD between class II genes (Dawkins *et al.* 1999), and in humans, there are particular hotspots for recombination within MHC regions and reduced recombination rates outside those hotspots (Cullen *et al.* 1997; Jeffreys, Kauppi & Neumann 2001). Selection may also act to prevent recombination from re-arranging beneficial allele combinations (Meyer & Thomson 2001), epistasis may vary between loci (Hedrick & Kim 2000) and selection acts on the group of genes within a haplotype (Navarro & Barton 2002). Divergent allele-specific selection pressures may therefore act on alleles at different loci in a single haplotype, which could effectively reduce, or cancel out any selection upon the haplotype. By generating information on how alleles are clustered on haplotypes helps to alleviate the one aspect of this challenge, by effectively reducing the number of loci to one whilst effectively representing the full haplotypic variation. Ultimately, selection can only act on functional variation, and this must be fully characterised in order to effectively detect and analyse modes of selection.

## 1.6 Gastrointestinal nematodes, the immune response and MHC in sheep

Gastro-intestinal nematodes (GIN) are a diverse group of helminth parasites that primarily infect the sheep abomasum and intestines and are a major contributor to costly pathogen associated morbidity and mortality in sheep (Nieuwhof & Bishop 2005; Sutherland & Scott 2009). GIN generate a range of pathologies due either directly to the activities of the worms, or indirectly through opportunistic bacterial infection or as a result of the immune response mounted by the host (Lawrence, Kennedy & Garside 2013). Larvae or eggs are ingested from the pasture, the parasites then develop into adult worms within the host gastrointestinal tract where they reproduce and the eggs are excreted in the host faeces. The acquired, or adaptive, immune response plays an important role during GIN infection, during which dendritic cells process and display parasite peptides on their MHC molecules, activate T cells either locally or by migrating to the lymph nodes (Lee *et al.* 2011). T helper 2 ($T_H2$) cytokine responses, which are essentially antibody-mediated responses, are typically induced by GIN infections, although T helper 1 ($T_H1$) cytokine responses, cell-mediated and inflammatory responses, can also occur (Maizels & Yazdanbakhsh 2003). Resistance to GIN typically manifests as reduced worm burden, worm size and worm fecundity (Lee *et al.* 2011).

The MHC plays an important role in mounting an immune response to GIN, and associations between MHC genes and disease in sheep have received much attention in an attempt to understand the processes influencing disease, as well as to develop marker-assisted breeding for resistance (Dukkipati *et al.* 2006b; Stear *et al.* 2007; Lee *et al.* 2011). Table 1.1 summarises association studies that have been carried out in domestic sheep *O. aries* on a variety of disease measures, including GIN burden measures such as faecal egg counts (FEC), viral diseases and bacterial infections, as well as immune response and fitness measures. All studies identified associations between MHC class IIa genotypes and traits of interest, although not necessarily for all traits studied, suggesting that the class IIa genes are involved in resistance or susceptibility to disease. Few studies have investigated the associations between class I and disease, and most have used class I antigen serotypes rather than sequence based methods (Dukkipati *et al.* 2006b), possibly due to the technical challenges in genotyping this region (see above). However, Figueroa Castillo *et al.* (2011) did identify an association between an allele at the class I linked microsatellite OMHC1 and reduced FEC and blood eosinophil counts. No such association with FEC was identified in a previous study in Soay sheep (Paterson, Wilson & Pemberton 1998). Few of these association studies in sheep have considered more than a single locus within an MHC class, and none characterised the diversity across the MHC class. Recently, the class II haplotypes were investigated in Texel sheep (Ali *et al.* 2016), though how these haplotypes relate to disease or fitness is unknown.

Table 1.1. MHC class II association studies in O. aries.

| Breed | Locus | Genotyping method | Age groups | Traits | Result | Reference |
|---|---|---|---|---|---|---|
| Broomfield Corriedale | DRA, DQA1, DQA2, DQB | RFLP | Lambs | Footrot (*Dichelobacter nodosus*) | Association between haplotype and resistance | (Escayg, Hickford & Bullock 1997) |
| Columbia, Polypay, Rambouillet | DRB1 | SBG | Ewes | Ovine progressive pneumonia virus | Two alleles associated with reduced virus levels, as were particular amino acids within allelic sequences. | (Herrmann-Hoesing *et al.* 2008b) |
| German Mutton Merino, Merino, Blackheaded Mutton | DQA2 | Unknown | Ewes | Footrot (*Dichelobacter nodosus*) | Presence of DQA2-like sequence associated with reduced footrot | (Ennen *et al.* 2009) |
| Latxa | DRB1 | SBG | Adults | Ovine pulmonary adenocarcinoma and Maedi-Visni disease (viral) | Alleles associated with susceptibility and resistance to diseases | (Larruskain *et al.* 2010) |
| Mehraban & Lori-Bakhtiari | DQA2 | SSCP & direct sequencing | Ewes | Lamb weights, milk production and serum protein measures | Genotypes & SNPs associated with milk and serum protein measures | (Dehkordi & Zamani 2015) |
| Merino | DRB1 | Microsatellites | Lambs (4 months) | Antibody responses to nematodes | Alleles associated with resistance | (Outteridge *et al.* 1996) |
| NZ Merino, SA Meat Merino, Polwarth and Corriedale | DQA1 | SSCP | Lambs (male; 4, 9 months) | FEC | DQA1*null not associated with FEC overall, but DQA1*null genotype associated with reduced FEC at 4 months in SA Meat Merino | (Forrest *et al.* 2010) |
| Perendale | DQA1 expression | cDNA expression array; Sanger sequencing | Lambs | GIN resistant vs susceptible lines | DQA1*null associated with susceptible line, but probably due to linkage with other alleles, | (Keane *et al.* 2007) |
| Polish Heath | DRB1 | Microsatellite (OLADRB) | Lambs (3-6 months); ewes (2-5 years) | FEC | Genotypes associated with FEC count | (Charon *et al.* 2002) |
| Polish Heath | DRB1 | Microsatellite (OLADRB) | Lambs (6, 12, 24 months) | Immune cell counts, Somatic cell count (mammary gland status) | One allele associated with both increased immune cell and somatic cell counts | (Swiderek *et al.* 2006) |

| Breed | Locus | Genotyping method | Age groups | Traits | Result | Reference |
|---|---|---|---|---|---|---|
| Scottish Blackface | DRB1 | Microsatellite | Various | FEC | Specific allele associated with reduced FEC | (Buitkamp & Epplen 1996) |
| Scottish Blackface | DRB1 | Microsatellite | Various | FEC | Common allele associated with reduction in FEC in lambs | (Schwaiger *et al.* 1995) |
| Scottish Blackface | DRB1 | Microsatellite | Lambs (4 month) | FEC | Allele likely to be associated with reduced FEC | (Stear *et al.* 1996) |
| Scottish Blackface | DRB1 | Microsatellite | Lambs | FEC | Allele associated with reduced FEC | (Stear, Innocent & Buitkamp 2005) |
| Scottish Blackface | Unkonwn | Microsatellite | Lambs (4-6 months) | FEC; Plasma anti-T. circ IgA | QTL on chromosome 20 within MHC region identified for both traits | (Davies *et al.* 2006) |
| Scottish Blackface, Sarda x Lacaune, Martinik Black-Belly x Romane | SNP | | Various | FEC | QTL on chromosome 20 for Sarda x Lacaune population only | (Riggio *et al.* 2014) |
| Soay | DRB1, DRB pseudogene, Class I | Microsatellite (OLADRB, OLADRBps, OMHC1) | Lambs (4 months), yearlings | FEC over-winter survival | Different DRB1 alleles each associated with reduced survival and increased FEC in lambs or with increased survival and reduced FEC in yearlings | (Paterson 1998) |
| Suffolk | DRB1 | SBG | Lambs | Worm burden, FEC, abomasal mast cell counts, immunoglobulin measures, haematological measures | DRB1*1101 carriers had reduced burden, higher mast cells and plasma platelet counts | (Hassan *et al.* 2011) |
| Suffolk & Texel | DRB1 | SBG | Lambs (2 months) | FEC | One allele associated with reduced and two with increased FEC in Suffolk but not Texel sheep | (Sayers *et al.* 2004) |

## 1.7 Soay sheep study system

The Soay sheep are descendants of an early domestication of sheep, *O. aries*, that have been living feral and virtually unmanaged on the remote island of Soay for thousands of years, and have continued to remain an unmanaged, wild population since the translocation of 107 sheep to the island of Hirta in 1932 (Clutton-Brock, Pemberton & Coulson 2004b) (Figure 1.5a). The Village Bay population (Figure 1.5b), which comprises between a quarter and a third of the population on Hirta (Clutton-Brock *et al.* 2004a), has been the focus of intensive study since 1985. The sheep are individually tagged as newborn lambs (Figure 1.5c), allowing the collection of life history data at an individual scale, as well as accurate population level data. Data and samples collected on newborns include weight, ear clips for genetic analysis and blood for analysis of immune measures. During August, the sheep are rounded up and morphometric measures, faecal samples for parasitological measures, and blood samples for immune measures are taken. Vegetation and climatic data are collected throughout the year. All breeding attempts and mortality is monitored, which along with a genetic pedigree based on hundreds of SNP loci, provide complete life history information for most individuals.

This long-term study provides a unique study system in which to investigate evolutionary and ecological questions, in part due to its relative simplicity compared to most ecological systems. Situated on a remote island, the Soay sheep population is effectively closed to migration. Predation is extremely limited and has virtually no impact on the population (Clutton-Brock *et al.* 2004b), and there are no food competitors (Crawley *et al.* 2004). The population undergoes large fluctuations in population density (Boyd *et al.* 1964), typified by high winter mortality following several years of population growth (Clutton-Brock *et al.* 2004a). Population dynamics in the Soay sheep are driven primarily by climate (wet, stormy winters increase mortality) and food limitation due to high population density (Clutton-Brock *et al.* 2004a). Furthermore, fluctuations in the demography of the population varies independently of the population size, and high post-winter mortality is more likely to occur under poor conditions when the population contains higher proportions of vulnerable individuals, lambs, males and older females (Coulson *et al.* 2001).

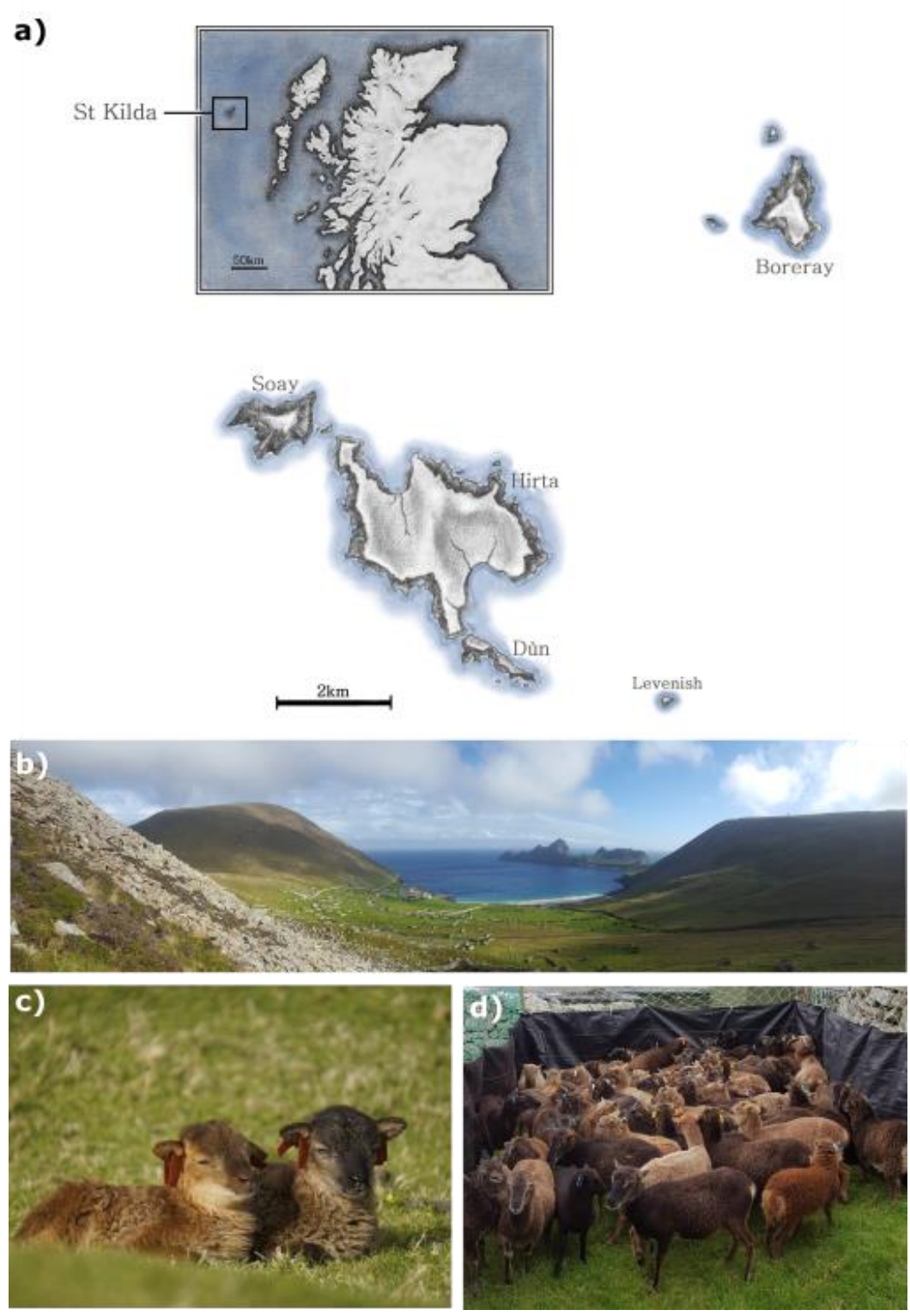



Figure 1.5. The Soay sheep are studied on the island of Hirta, St. Kilda (a), which is located off the west coast of Scotland. The individuals which live within the skyline of Village Bay (b) on the south-east of the island are tagged shortly after birth (c), and caught in August (d) when a number of phenotypic measurements are taken.

The parasite community within the Soay sheep is relatively well understood (Table 1.2; Wilson *et al.* 2004). Being an island system, parasite diversity is limited by their ability to reach the island, lack of vectors or intermediate hosts and ability to live endemically in a small host population (Wilson *et al.* 2004). The parasite community is therefore composed primarily of species with direct life-cycles and limited pathogenicity. Annual measures of individual sheep parasite burdens are collected using the indirect measure of faecal egg count (FEC – nematodes, tapeworms), faecal oocyte count (FOC – protozoa) and absolute counts (flies) (Table 1.2).

Table 1.2. Parasite species known to infect Soay sheep. During the summer catch, measures of parasite burden are collected as detailed. Counts are combined for strongyle species and Nematodirus species.

| Taxon | Species name | Yearly measures |
|---|---|---|
| Nematodes | *Capillaria longipes* | FEC |
| | *Trichuris ovis* | FEC |
| | *Nematodirus battus* | FEC *Nematodirus* |
| | *Nematodirus filicollis* | |
| | *Nematodirus helvetianus* | |
| | *Bunostomum trigonocephalum* | FEC strongyles |
| | *Chabertia ovina* | |
| | *Stongyloides papillosus* | |
| | *Teladorsagia circumcincta* | |
| | *Trichostrongylus axei* | |
| | *Trichostrongylus virtinus* | |
| | *Dictyocaulus filaria* | Not measured |
| | *Muellerius capillaris* | Not measured |
| Tapeworms | *Monezia expansa* | Faecal presence/absence |
| | *Taenia hydatigena* | Not measured |
| Lice | *Damalina ovis* | Not measured |
| Flies | *Melophagus ovinus* | Absolute count |
| Bacteria | *Dermatophilus congolensis* | Not measured |
| Protozoa | *Cryptosporidium parvum* | Not measured |
| | *Eimeria* species x11 | FOC Coccidea |
| | *Giardia duodenalis* | Not measured |

Source: Wilson *et al.* 2004

Parasites are known to be an important selective force in the system. Individuals with a greater parasite burden are more likely to die earlier in a population crash (Gulland 1992); survival was greater in female lambs and male yearlings treated with an anthelminthic bolus prior to the 1992 crash than untreated controls (Gulland *et al.* 1993), and fitness is negatively correlated with FEC in lambs (Hayward *et al.* 2011).

Although winter climate and lack of forage due to high sheep density are the primary drivers of the population crashes that occur in the Soay sheep (Clutton-Brock *et al.* 2004a), parasite resistance is likely to be a key factor in determining which individuals survive high mortality winters and must, therefore, be under strong selection (Gulland 1992). Resistance to GIN, measured as FEC, is heritable in lambs (0.26 ± 0.12 Beraldi *et al.* 2007) and has a low heritability in adults (Females: 0.0 – 0.05 ± 0.02, Males: 0.0 – 0.12 ± 0.03 (Beraldi *et al.* 2007; Robinson *et al.* 2009)), demonstrating that there is a genetic basis to GIN resistance.

Discriminating between modes of selection acting on the MHC will only be possible within a study system which is simple but not pathogen depauperate, and for which genotypic and phenotypic data are available at both sufficient temporal and spatial scales (Spurgin & Richardson 2010). The Soay sheep data are therefore well suited for such analysis. The population is spatially heterogeneous both genetically (Coltman, Pilkington & Pemberton 2003; Charbonnel & Pemberton 2005) and parasitologically (Wilson *et al.* 2004), with parasitological data already existing over a 30 year period. Finally, the extensive individual-based life-history data and immunological measures available will enable the relationships between MHC genotype, fitness and immune function to be investigated. Associations with fitness measures, including lifetime reproductive success and survival, may provide insight into the impact of MHC haplotype on individual fitness. Long-term data on parasitaemia is available as an indirect measure of parasite burden, FEC. FEC has been shown to correlate with worm burden within this study population (Wilson *et al.* 2004) and therefore should be valuable in assessing association between MHC haplotype and parasite resistance. Furthermore, a variety of immune measures have been collected (Graham *et al.* 2010; Nussey *et al.* 2014) which can be investigated in relation to MHC variation and fitness.

### 1.7.1 MHC in the Soay sheep

Associations between MHC genotype, parasitaemia and fitness have been investigated in the Soay sheep previously, using linked microsatellite markers (Paterson 1996, 1998; Paterson *et al.* 1998), but in light of the foregoing, they should perhaps be treated as preliminary. Alleles of the OLADRB microsatellite, located adjacent to the *Ovar-DRB1* exon 2, were found to be associated with survival in lambs and yearlings, and parasite resistance (FEC) in lambs (Paterson *et al.* 1998). OLADRB microsatellite allele frequencies had a more even distribution than expected under neutrality, suggestive of

balancing selection (Paterson 1998). Further investigation of the *DRB1* locus itself showed evidence that selection has operated over the history of the population, as the non-synonymous substitution rate was greater than for synonymous substitutions at the PBR (Paterson 1998). Evidence is therefore strongly suggestive that balancing selection operates at *DRB1* in the Soay sheep. Furthermore, Charbonnel and Pemberton (2005) found that OLADRB microsatellite alleles varied spatially and temporally within the Soay sheep (independently of neutral loci), indicative of fluctuating selection. No evidence has been found to suggest that MHC disassortative mating operates within the Soay sheep (Paterson & Pemberton 1997). The strengths of conclusions drawn from these studies, however, are limited as they suffer from a number of the limitations discussed in section 1.5. Firstly, microsatellite genotypes rather than expressed locus genotypes were assessed (Paterson *et al.* 1998; Charbonnel & Pemberton 2005). Paterson et al. (1998) genotyped three microsatellite loci, two class II associated with *DRB1* (OLADRB and OLADRBps) and one class I (OMHCI); however the association between OMHCI and any expressed variation was unknown and may not reflect functional MHC class I variation (Paterson 1998). Thus, only a single class II locus has been effectively analysed.

## 1.8 Thesis outline

In this thesis, I aim to use advances in genotyping methodologies for the MHC class IIa region to generate haplotype data for the Soay sheep population to investigate how balancing selection acts on this region using the extensive phenotypic data collected over 30 years. The following sections provide a brief overview and the aims of each chapter.

### 1.8.1 Chapter 2: Major histocompatibility complex class IIa haplotypes in the Soay sheep population

In chapter 2, I use genotyping-by-sequencing methods developed for domestic sheep to characterise seven loci within the class IIa region. These loci are *DRB1, DQA1, DQA2, DQA2-like, DQB1, DQB2* and *DQB2-like*. Using a strategic genotyping strategy, I genotyped *DRB1* in 118 sheep randomly selected from four cohorts to identify homozygous sheep, which were subsequently genotyped at the remaining six loci. A test sample of 94 individuals was then typed at the *DRB1* and *DQA* loci to determine if

the identified haplotypes held true in a sample not selected on the basis of homozygosity.

### 1.8.2 Chapter 3: Selection and testing of a custom SNP panel for rapid determination of Class IIa haplotypes in large numbers of Soay sheep

In chapter 3, I make use of the haplotype information from chapter 2 to develop a SNP based genotyping system for the class IIa loci. I use previously collected genotype data from the Ovine Infinium HD Beadchip (with 8420 variable SNPs on chromosome 20) to identify a linkage region around the class IIa genes. From the subset of SNPs within the identified linkage region, I identify a panel of 11 SNPs that can be used to infer the class II haplotype, and add two intragenic SNPs which are capable of detecting a recombinant haplotype identified by the Ovine Infininium HD Beadchip analyses. Finally, I test this panel of 13 SNPs on a sample of the Soay population using Kompetitive Allele Specific PCR. I discuss the success of the KASP SNP panel, and the ability of the intragenic SNPs to detect the recombinant haplotype.

### 1.8.3 Chapter 4: Kompetitive Allele-Specific PCR (KASP) genotype analysis of MHC class IIa haplotype diversity in a large sample of Soay sheep

In chapter 4, I genotype a large proportion of the Soay sheep population born between 1985 and 2012 using the KASP SNP panel developed in chapter 3. I then discuss the quality control measures applied to assess the genotyping quality through assessment of individual genotyping rate, consistency of genotypes for a repeatedly genotyped control sample, and use of Mendelian inheritance checks. I also discuss identifying the source of the recombination event (identified in chapter 3) and its inheritance throughout the individual's descendants. Finally, I investigate whether the data are in Hardy-Weinberg Equilibrium, whether this changes over time or at different life-history stages, and whether diplotype frequencies vary over time.

### 1.8.4 Chapter 5: Associations between haplotypes and health-related phenotypes

Chapter 5 investigates whether haplotypes are associated with trait values and begins to disentangle the mechanisms of balancing selection by asking whether class IIa heterozygosity or haplotypes are associated with health-related phenotypes. The health-related phenotypes tested are August weight, immunoglobulin titers for anti-

*Teladorsagia circumcincta* IgA, IgE and IgG, and FEC. First, I test for associations between heterozygosity status and trait variation, and then for associations between haplotype presence or absence, or haplotype dosage (absent, one copy, two copies) and traits using mixed effects models. Because related individuals are included within this data, I also conduct these tests using animal models to account for similarity in trait values due to relatedness.

# 2. Major histocompatibility complex class IIa haplotypes in the Soay sheep population

## 2.1 Abstract

Rival theories for the evolution and maintenance of exceptional diversity at the MHC are difficult to discriminate, especially when the region is poorly characterised. The mode of selection would be best studied using well-characterised haplotypes bearing combinations of alleles at different adjacent expressed loci. A structured sequence-based-genotyping strategy was used to identify alleles at seven MHC class IIa loci in a sample of 118 Soay sheep selected at random from four cohorts spanning 15 years of the St. Kilda study. *DRB1* is the best characterised and most polymorphic locus in sheep, and this was characterised first in all 118 sheep. Then, the *DQA* and *DQB* loci were characterised using *DRB1* homozygotes (from gDNA or cDNA). Eight haplotypes were identified within the Soay sheep population, with some alleles shared across haplotypes. A test sample of 94 further individuals typed at *DRB1* and *DQA* loci found no exceptions to the eight identified haplotypes, and haplotype homozygosity rate of 21.3 %. Both *DQ1/DQ2* and *DQ2/DQ2-like* haplotype configurations were identified, as well as a haplotype carrying a *DQB1* and two *DQB2* alleles.

## 2.2 Introduction

The major histocompatibility complex (MHC) is a genomic region containing highly polymorphic genes which encode cell surface proteins involved in the presentation of pathogen-derived peptides to T cells, enabling an immune response (Klein 1986). Pathogen-mediated balancing selection is now commonly recognised as the major driver of genetic variation at the MHC, but the different, though not mutually exclusive, mechanisms which could be operating (heterozygote advantage, frequency dependent selection and fluctuating selection) are difficult to differentiate because their predicted effects on MHC dynamics are very similar (Spurgin & Richardson 2010). For example, rare alleles which are advantageous under frequency-dependent selection most commonly occur in heterozygotes, which are advantageous under heterozygous advantage (Apanius *et al.* 1997). The challenge of disentangling the effects of different modes of balancing selection is made worse if MHC genotyping methods are imprecise, for example, if alleles cannot be attributed to loci.

The class IIa loci exhibit high linkage disequilibrium (LD) (Lee *et al.* 2012), which can make it difficult to disentangle the effects of alleles and haplotypes (Stear *et al.* 2007; Ali *et al.* 2016). Due to the technical challenges of genotyping the MHC in non-model species, many studies take advantage of this linkage disequilibrium between loci and genotype a single locus, assuming that this is representative of the full haplotypic diversity (e.g. Babik *et al.* 2005; Harf & Sommer 2005; Bollmer *et al.* 2007; Biedrzycka *et al.* 2011; Gelasakis *et al.* 2013; Kamath *et al.* 2014; see Appendix A). However, despite high LD, selection favouring particular alleles could mean that some alleles at specific loci are identical across different haplotypes (de Bakker *et al.* 2006; Traherne *et al.* 2006). For example, in sheep, an allele at one locus can be found in combination with multiple alleles at other loci (Hickford *et al.* 2007; Ballingall *et al.* 2015; Ali *et al.* 2016). Genotyping only a single locus may disguise variation at other loci, reducing the power to detect differing selection pressures. This highlights the need to characterise haplotypic variation when studying the MHC.

The highly complex and polymorphic nature of the MHC region has made it a focus of many studies in immunology and evolution. However, these features make it difficult to genotype individual loci. Over evolutionary time, mammalian MHC loci can be viewed as going through a birth and death process (Nei et al. 1997). New MHC loci are thought to be created through gene duplication events, resulting in multiple loci harbouring similar or identical alleles. Some are lost through decay, producing pseudogenes and gene fragments which are characteristic of mammalian MHC regions. Additionally, allelic diversity within MHC loci is high, and selection appears to favour the maintenance of numerous divergent alleles within a population. Locus-specific assays can be challenging to develop, as multiple loci and pseudogenes may co-amplify with primers that are too generic, and allelic dropout may occur with primers that are too specific (Babik et al. 2009). Ideally, a study of selection on the MHC would overcome the problem of accurate genotyping of multiple loci, identify the haplotypes present and study selection on haplotypes or diplotypes in a large sample of individuals.

The unmanaged but intensively studied Soay sheep (*O. aries*) population on the island of Hirta, in the St. Kilda archipelago, Scotland, presents an excellent opportunity to study MHC selection in a large mammal. Since 1985 individual-level data on phenotypic traits, parasite burdens and life-history and have been collected for many hundreds of individuals (Clutton-Brock *et al.* 2004b). Gastrointestinal nematodes are important parasites of the Soay sheep, and individuals with high strongylid faecal eggs counts in

summer are less likely to survive the subsequent winter (Wilson *et al.* 2004; Hayward *et al.* 2011, 2014a). Moreover, the Hirta population of Soay sheep originated from 107 animals which were translocated from the neighbouring island of Soay in 1932 (Clutton-Brock *et al.* 2004b). These 107 sheep included 20 rams, 44 ewes, 21 ewe lambs and 22 castrated ram lambs (Boyd 1953; Boyd *et al.* 1964). The founding population thus consisted of 85 animals which could potentially contribute to the gene pool, with an estimated effective population size ($N_e$) of approximately 61.2 individuals (calculated as

$$N_e = \frac{(4 * Number\ of\ males * Number\ of\ females)}{(Number\ of\ males * Number\ of\ females)}$$

to account for a smaller number of breeding males than females Charlesworth 2009). Additionally, the original population on the island of Soay is itself is only a few hundred individuals at any one time (Clutton-Brock *et al.* 2004b). MHC variation within the study population is therefore expected to be more limited when compared with larger populations experiencing immigration. Fully characterising the MHC class IIa haplotype variation within the Soay population might, therefore, be feasible.

The two major families of antigen-presenting MHC molecules are class I, which present endogenous antigens (primarily from intracellular pathogens) to CD8 cells, and class II which present exogenous antigens (primarily from extracellular pathogens) to CD4+ T cells (Klein 1986). Unusually, within ruminants, the class II region is split into two distinct subregions, class IIa and IIb (Andersson *et al.* 1988; van Eijk *et al.* 1995). The classical class IIa loci, which have previously been associated with parasite resistance (see Lee *et al.* 2011), are clustered in the class IIa region. These include the highly polymorphic *DRB1*, *DQA* and *DQB* loci, and the less polymorphic *DRA* (Ballingall *et al.* 2010). Duplicated pairs of *DQA* and *DQB* loci have been identified in domestic sheep, *Ovis aries* (Scott *et al.* 1987; Wright & Ballingall 1994; Ballingall *et al.* 2015, 2017 in press). Three types of *DQA* alleles (*DQA1*, *DQA2* and *DQA2-like*; (Ballingall *et al.* 2015) and *DQB* alleles (*DQB1*, *DQB2* and *DQB2-like*) (Ballingall *et al.* 2017 in press) have been identified. Whilst *DQA1* and *DQA2* are known to be different loci, as are *DQB1* and *DQB2,* origins of the *DQA2-like* and *DQB2-like* alleles are less well understood. Typically, *DQA2-like* and *DQB2-like* alleles are found on haplotypes in conjunction with *DQA2* and *DQB2*, on which the *DQA1* and *DQB1* loci are absent. Therefore, the two typical haplotype configurations are *DQA1+DQA2* with *DQB1+DQB2* and *DQA2+DQA2-like* with *DQB2+DQB2-like*. Whether the *DQA2-like* and *DQB2-like* alleles represent independent

loci or are simply divergent alleles at the *DQA1* and *DQB1* loci is still not clear (Ballingall *et al.* 2015, 2017 in press). A recent study by Ali *et al.* (2016) identified haplotypes with all three allele types, which would suggest that the *DQA2-like* and *DQB2-like* alleles are derived from independent loci.

A previous study of MHC variation in Soay sheep using the OLADRB microsatellite located in the second intron of the class IIa locus *DRB1* found evidence for selection acting on this region (Paterson *et al.* 1998). Paterson *et al.* (1998), found that a specific OLADRB allele was associated with increased yearling survival and parasite resistance, while another allele was associated with reduced survival and parasite resistance in lambs, suggesting that different MHC alleles may confer differing advantages at different life history stages. However, as outlined above, how well the single OLADRB microsatellite locus represents diversity across the MHC class IIa region of the Soay sheep is not clear. Since the Paterson *et al.* (1998) study, single locus genotyping methods targeting the polymorphic regions of the classical class IIa loci have been developed for domestic sheep. These include *DRB1* (Ballingall & Tassi 2010), *DQA* (Ballingall *et al.* 2015) and *DQB* (Ballingall *et al.* 2017 in press). In this study, my aims were (1) to genotype a sample of Soay sheep at the classical *DRB1* and *DQ* loci using sequence-based genotyping and (2) to define the class IIa haplotypes in the study population. Characterising the class IIa loci in this population will facilitate the development of a method to determine haplotypes for large numbers of individuals, enabling subsequent investigation of the evolutionary mechanism maintaining diversity within this region.

## 2.3 Methods

### 2.3.1 Study system

Monitoring of Soay sheep in the Village Bay area on Hirta has been carried out intensively since 1985 (Clutton-Brock *et al.* 2004b), including catching lambs in spring for weighing, ear-tagging and sampling for genetic analysis. Most sheep are also caught in August, when phenotypic measurements, faecal samples for strongyle egg counts and blood samples are taken. During the August catches of 2012 to 2014, aliquots of blood were collected into Tempus™ Blood RNA Tubes (ThermoFisher Scientific).

## 2.3.2 Sequence-based genotyping

### 2.3.2.1 *DRB1* genotyping

The *DRB1* locus is the best characterised class II locus in *O. aries*, and the Immuno Polymorphism Database (IPD-MHC - http://www.ebi.ac.uk/ipd/mhc/ovar/) contains over 100 alleles and a corresponding allelic nomenclature. Locus-specific primers and a sequence based genotyping method targeting the polymorphic second exon have previously been described (Ballingall & Tassi 2010). The primer pair 330_F and 329_R (Ballingall & Tassi 2010) was tested initially, as it generates full *DRB1* exon 2 sequences. However, due to a 1 bp deletion in the *DRB1*13:01* allele, the forward primer 455_F (a modification described in (Corbishley *et al.* 2016)) which sits downstream of the deletion was used in preference (Table 2.1). Between 27 and 31 sheep were randomly selected from each of four cohorts (1993, 1998, 2003, 2008) and genotyped to determine allelic diversity within the Soay population and to identify *DRB1* homozygous individuals for subsequent analysis of *DQ* diversity.

### 2.3.2.2 DQA genotyping

For each *DRB1* allele identified, four *DRB1* homozygous individuals were genotyped at the *DQA* loci (*DQA1*, *DQA2* and *DQA2-like*). *DRB1* homozygous individuals were selected, as they were more likely to be homozygous at DQ loci due to linkage disequilibrium. Allele (*DRB1*10:01*) was not identified in homozygous form and thus *DQA* and *DQB* haplotypes were determined from heterozygous individuals. Primers DQA1_F and DQA1_R were used to amplify *DQA1* and primers DQA2_F and DQA2_R were used to amplify *DQA2* and *DQA2-like* (Table 2.1). For individuals that were homozygous at the haplotype level and carried the *DQA2/DQA2-like* configuration, the *DQA1* primers were expected to fail to generate a PCR product and the *DQA2* primers were expected to generate a heterozygous product.

### 2.3.2.3 DQB genotyping

*DQB* loci (*DQB1*, *DQB2* and *DQB2-like*) were characterised in individuals which were both *DRB1* and *DQA* homozygous, and from heterozygous individuals for *DRB1*10:01*. Primers DQB-F and DQB1-R or DQB2-R were used to amplify the *DQB* loci (Table 2.1). *DQB* primers were not completely locus-specific and some cross-amplification was expected, depending upon the alleles present (Ballingall *et al.* 2017 in press).

#### 2.3.2.4 PCR amplification

PCR reactions were carried out in a final volume of 25 µL and contained 12.5 µL Promega GoTaq Green mastermix, 0.5 µM of each primer, approximately 25 ng genomic DNA and water. Cycling conditions were 94 °C for 5 mins, then 35 cycles of 94 °C for 30 or 60 s, 55-58 °C for 30 or 60 s, and 72 °C for 30 or 60 s (see Table 2.1), and a final extension of 72 °C for 5 mins. PCR products were checked using gel electrophoresis on a 1 % agarose gel.

#### 2.3.2.5 RT-PCR amplification of full length DQ transcripts

Full length transcripts amplified from RNA were required for previously unidentified alleles. For individuals for which Tempus™ Blood RNA Tubes were available (see Study system), *DRB1* and *DQA* loci were amplified from genomic DNA to identify homozygous individuals for RNA extraction. Total RNA was extracted from 3 mL of Tempus tube blood stored at -20 °C using the Tempus™ Spin RNA Isolation Kit (ThermoFisher Scientific) at half the recommended volumes. Reverse Transcription was carried out using the ImProm-II™ kit (Promega) with oligo$(dT)_{15}$ primers. RT-PCR reactions were carried out using the primer combinations described in Table 2.1. No single primer set amplified all alleles. As RNA was not available for any individuals homozygous for haplotype H, F/H heterozygous individuals were used instead.

#### 2.3.2.6 Cloning of multi-allelic PCR products

In order to phase novel alleles from heterozygous or multi-locus amplifications, they were cloned into the pGEM-T easy vector (Promega). Presence of the PCR insert was confirmed using colony PCR and plasmid DNA from 12 to 20 colonies was purified for sequencing using QIAprep Spin Miniprep Kit (Qiagen).

#### 2.3.2.7 Sequencing and sequence analysis

PCR products were purified using exonuclease I and Antarctic phosphatase, except *DQB2* PCR products which were gel-purified using the Macherey-Nagel Nucleospin Gel and PCR Clean Up kit, prior to Sanger sequencing. Alleles were sequenced using BigDye 3.1 chemistries on an AB 3730 genetic analyser.

Sequence analysis was carried out in Geneious 7.1.9. For *DRB1* sequences, heterozygous peaks were called using the Heterozygote Plugin and checked by eye. *DRB1* sequences were then compared to alleles from the IPD-MHC database using a custom BLAST. *DRB1* genotypes were called when known alleles accounted for all variants within the sequence.

*DQA* and *DQB* sequences were also analysed using the Heterozygote Plugin within Geneious 7.1.9 to call heterozygous peaks. Due to multi-locus amplification, some sites contained three or four peaks, which were called by eye. A custom BLAST was used to compare sequences to an appropriate *DQA* database (Ballingall *et al.* 2015) or *DQB* database (Ballingall *et al.* 2017 in press). Genotypes were only called when all variants were accounted for by known alleles. Following cloning and RT-PCR of unknown alleles, the custom BLAST was updated to include these. Uncalled sequences were then compared to the updated database and assigned temporary nomenclature (see Appendix B).

### 2.3.3 Phylogeny of DQB

Due to the lack of locus-specific DQB primers, loci could not be determined for all amplified alleles, and thus a phylogeny of the DQB sequences was generated in an attempt to determine their DQB locus of origin. The DQB alleles sequenced here were aligned using Clustal Omega (Sievers *et al.* 2011) with previously identified DQB sequences from *O. aries* (Ballingall *et al.* 2017 in press), *Bos taurus*, and *B. mutus*. The human DQB sequence (accession number M24364.1) was used as an outgroup. Model selection for MrBayes, implemented in Topali v2 (Milne *et al.* 2009), selected the K80 model of DNA substitution (Kimura 1980) with gamma distribution. The phylogeny was generated in Geneious v7.1.9 using MrBayes 3.2.6 (Huelsenbeck & Ronquist 2001) using 1,000,000 generations.

### 2.3.4 Pairwise amino acid differences

The total number of amino acid differences between haplotypes was calculated in MEGA 7.0.21 (Kumar, Stecher & Tamura 2016) from exon 2 of *DRB1*, *DQA1* or *DQA2-like*, *DQA2*, *DQB1* or *DQB2-like* and *DQB2*. Haplotype G was included twice with either HHDQB1N or HHDQB2B1 as the DQB2 allele. DQB sequences were unknown for haplotype D, and so the total number of amino acid differences for *DRB1*, *DQA1* or *DQA2-like*, *DQA2* only were calculated for this haplotype.

### 2.3.5 Validation of haplotypes

To confirm that the haplotypes characterised in homozygous individuals were consistent, *DRB1* and *DQA* loci were genotyped in 95 sheep of unknown MHC genotypes. These sheep were not first order relatives and were born between 1984 and 2010. These individuals were selected as they were genotyped on the Ovine Infinium

HD SNP BeadChip for a previous study (Johnston *et al.* 2016). These individuals represent the maximum genetic variation within the population for the given number of individuals (Johnston *et al.* 2016).

Table 2.1. Details of the primers used and their PCR conditions.

| Target Loci | DNA target | Primer name | Primer sequence | $T_A$ | PCR cycle times (secs) | Primer reference |
|---|---|---|---|---|---|---|
| *DRB1* | gDNA | 330_F | ATTAGCCTCYCCCCAGGAGKC | 55 | 30, 30, 30 | (Ballingall & Tassi 2010) |
| | | 455_F | TATCCCGTCTCTGCAGCACATTTC | 58 | | (Corbishley *et al.* 2016) |
| | | 329_R | CACCCCCGCGCTCACCTCGCCGC | 55-58 | | (Ballingall & Tassi 2010) |
| *DQA1* | gDNA | *DQA1*_F | ACCTGACTCACCTGACCACA | 55 | 60, 60, 60 | (Ballingall *et al.* 2015) |
| | | *DQA1*_R | AACACATACTGTTGGTAGCAGCA | 55 | | |
| *DQA2* and 2-like | gDNA | *DQA2*_F | ACTACCAATCTCATGGTCCCTCT | 58 | | |
| | | *DQA2*_R | GGAGTAGAATGGTGGACACTTACC | 58 | | |
| *DQA1* and 2 | cDNA | 244_F | GCTGAGMCCACCTTGAGAASAG | 55 | 60, 60, 60 | (Ballingall *et al.* 2015) |
| | | 241_R | TGAGATGATACAGCATCTTAAGTCC | 55 | | |
| *DQA1*, 2 and 2-like | cDNA | 348_F | GAGGATGGTCCTGAACAGAGC | 55 | | |
| | | 357_R | GAGGAGGGCAGAAGAAGAAAA | 55 | | |
| *DQB* | gDNA | *DQB*_F | CCCCGCAGAGGATTTCSTG | 58-60 | 30, 30, 30 | (Ballingall *et al.* 2017 in press) |
| *DQB1* | gDNA | *DQB1*_R | CGGCACTCACCTCGCCGCTGC | 60 | | |
| *DQB2* and 2-like | gDNA | *DQB2*_R | ACGCTCACCTCGCCGCTGCC | 58 | | |
| *DQB1* and 2 | cDNA | 245_F | TGGGTGTTGACTACCATTAST | 55 | 60, 60, 60 | (Ballingall *et al.* 2017 in press) |
| | | 248_F | ACGCASSYATTAYAGAAGAGC | 55 | | |
| *DQB2-like* | cDNA | 392_F | ATTAGTTGTTCCTTTTTTCTC | 55 | | |
| | | 395_R | AAAATATCCTCAGGAGTCAGC | 55 | | |
| | | 401_R | CAAGAACACGCAGCTATTACA | 55 | | |

# 2.4 Results

## 2.4.1 Sequence-based genotyping

### *2.4.1.1 DRB1*

Six *DRB1* alleles were identified among the 118 individuals from the four cohorts born in 1993, 1998, 2003 and 2008. All six alleles were represented in the IPD-MHC database and each allele was assigned to an individual haplotype (A-F, Table 2.2). Allele frequencies are shown in Figure 2.1. Alleles *01:01, *01:02 and *22:01 appear at the highest frequencies in all four cohorts while *03:02, *10:01 and *1301 are consistently the three lowest frequency alleles. In total, 33 (27.5 %) *DRB1* homozygous sheep were identified, with five of the six *DRB1* alleles represented by a minimum of three homozygous individuals. The allele at the lowest frequency, *DRB1**10:01 (Figure 2.1), was only observed in six heterozygous individuals.

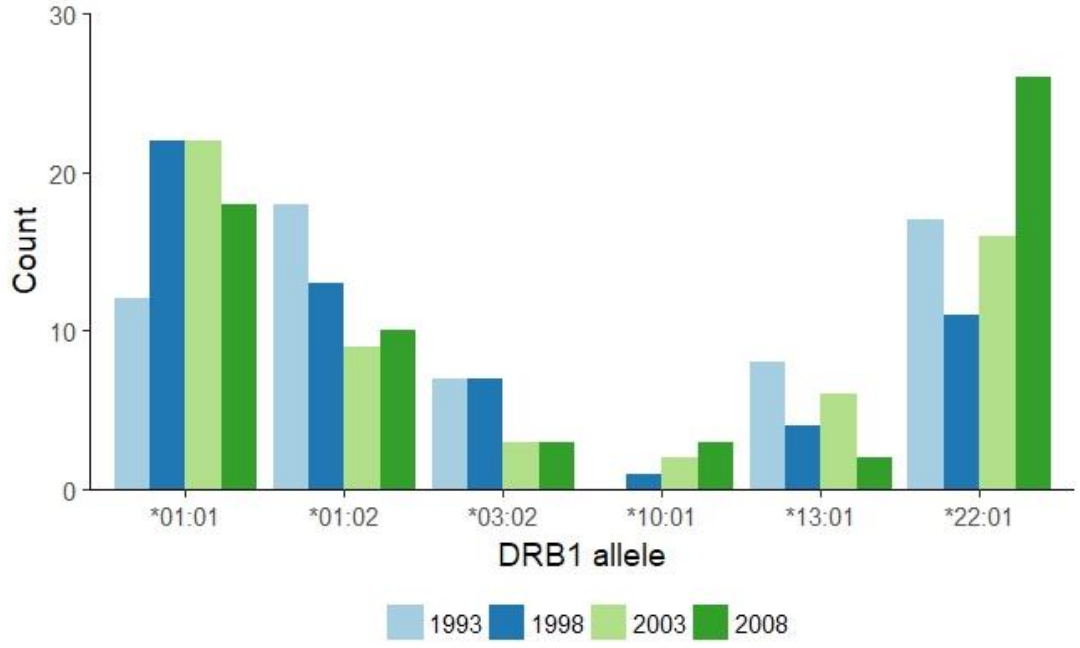



Figure 2.1. Frequency of DRB1 alleles in four cohorts of Soay sheep.

Table 2.2. MHC class IIa haplotypes identified in the Soay sheep across *DRB1*, *DQA* and *DQB* loci. A * preceding the allele name indicates nomenclature follows IPD guidelines (Marsh *et al.* 2010) - only *DRB1* alleles were included in the IPD for *O. aries* at the time or writing. *DQ* alleles which have not been previously described by Ballingall *et al.* (2015) or Ballingall (in press), or were identified by this study and assigned IPD nomenclature within Ballingall (in press) are highlighted in bold and were named according to the GenBank sequence which they match or were assigned temporary nomenclature (NewX). *DQB* loci were assigned according to phylogenetic cluster (Figure 2.4).

| Haplotype | *DRB1* | *DQA1* | *DQA2* | *DQA2-like* | *DQB1* | *DQB2* | *DQB2-like* |
|---|---|---|---|---|---|---|---|
| A | *01:01 | | NewA | NewA | | *09:01 | *03:01 |
| B | *01:02 | *03:01:01 | *01:01:01 | | *02:01:01 | *04:01:01 | |
| C | *03:02 | Z28420 | *07:01:01 | | AJ238933 | AJ238945 | |
| D | *10:01 | *03:01:01 | *02:01:01 | | Unknown sequence | Unknown sequence | |
| E | *13:01 | | *01:02:01 | *01:01:01 | | *10:01 | *01:01:01 |
| F | *22:01 | NewF | NewF | | NewF[a] | AJ23941 | |
| G | *01:01 | | *01:02:01 | *01:01:01 | | *09:01 & *12:01 | *01:01:01 |
| H | *22:01 | NewH | NewH | | NewH | *11:01 | |

[a] Phylogenetic relationship of this allele remains unclear as both haplotype F alleles fell within the *DQB2* cluster (see Figure 2.4), although NewF was grouped with another *DQB1* allele within the *DQB2* cluster. The most parsimonious assumption is that NewF is *DQB1*.

### 2.4.1.2 DQA loci

*DQA* genotyping of *DRB1* homozygous individuals and *DRB1*10:01* heterozygotes provided evidence for an additional two haplotypes (G and H) (Table 2.2). The *DRB1*01:01* and *DRB1*22:01* alleles were each associated with two different *DQA* haplotypes. *DQA1* primers failed to amplify a product in three haplotypes, (*DQA1* null haplotypes, A, E and G). In each of these haplotypes a *DQA2-like* allele was identified.

Six *DQA* alleles did not match full length sequences described in (Ballingall *et al.* 2015), and only one of these (*DQA1**Z28420 on haplotype C) matched a sequence on GenBank (accession number Z28420). Full length transcripts (768 bp) were generated for six of these *DQA* alleles (Table 2.2). *DQA1* could not be amplified from cDNA for haplotype C and nomenclature was not assigned. An alignment of the predicted DQα proteins is shown in Figure 2.2.

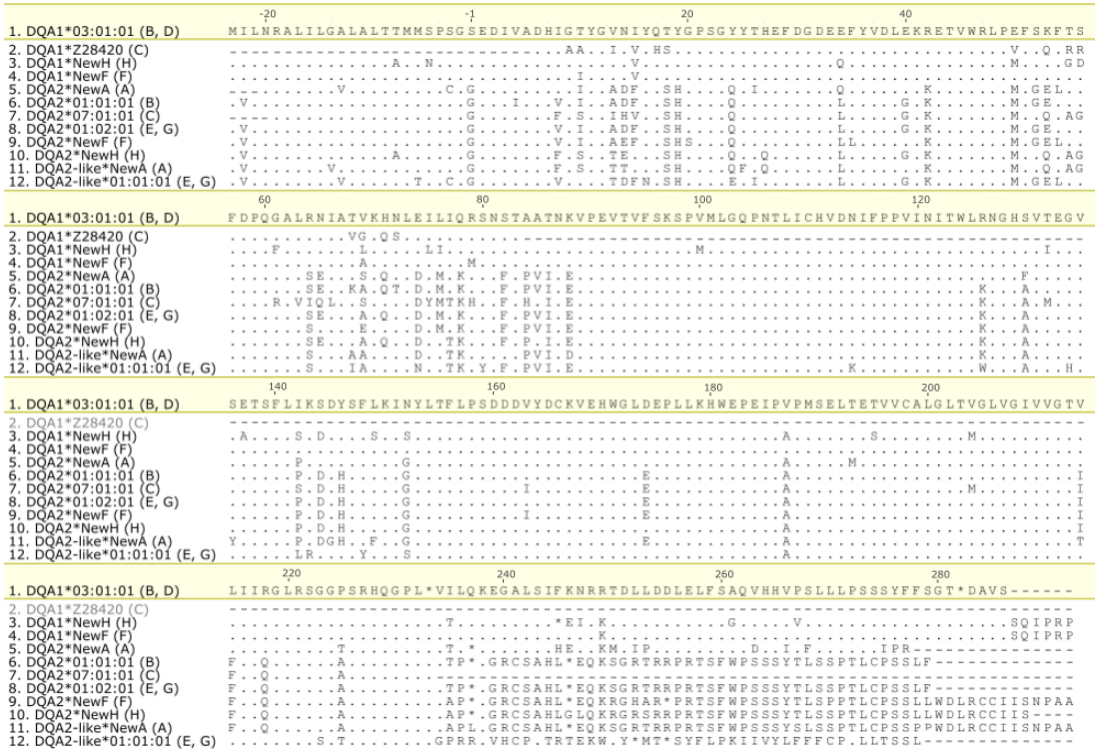



Figure 2.2. Predicted *DQα* sequences identified in Soay sheep. Sequences are numbered from the first amino acid of the mature protein. Dots (.) indicate identity to sequence 1 and dashes (-) indicate missing sequence. The haplotypes on which each allele is present is shown in brackets after the allele name.

#### 2.4.1.3 *DQB* loci

Both gDNA and cDNA *DQB* primer pairs exhibited varying degrees of cross-locus amplification, dependent upon the alleles present (Table 2.3). The *DQB2-like* cDNA primers did not amplify any alleles from haplotypes known to carry a *DQB2-like* allele, likely due to polymorphisms at the primer binding regions but perhaps due to poor quality RNA samples.

A single allele at each of the loci matched alleles in Ballingall *et al.* (in press), and a further two alleles at each of *DQB1* and *DQB2* matched alleles in GenBank (Table 2.2, Table 2.3). The remaining alleles were novel and were assigned temporary nomenclature (Table 2.2). Full-length sequences for the two novel alleles on haplotypes E and H could not be generated and full-length transcripts failed to amplify for the other five alleles in haplotypes C, F and H as shown in Table 2.3. Three DQB alleles were identified from haplotype G, two from gDNA only and one from cDNA only (Table 2.3). An alignment of the predicted DQα proteins is shown in Figure 2.4.

*DQB* alleles on haplotype D, for which only heterozygous animals were available, could not be resolved. Sequences obtained using gDNA *DQB1* and *DQB2* primers were both heterozygous for the known allele from the alternative haplotype and a novel allele. Cloning *DQB* transcripts from an animal heterozygous for haplotypes B and D did not provide alleles from haplotype D, suggesting that the cDNA primers failed to amplify haplotype D alleles.

Table 2.3. *DQB* alleles identified for each class IIa haplotype. Allele name is shown as included in the *DQB* phylogeny. Success of the amplification from genomic DNA and cDNA is indicated. Numbers in brackets indicate which reverse primer (Table 2.1) was used in conjunction with the common forward primer. Sequences identified from gDNA which matched previously identified alleles were not sequenced from cDNA (NA). *DQB* locus assignment is based on the phylogenetic cluster the allele was located in, and the length of the allelic sequence included in the phylogeny is shown.

| Class IIa haplotype | Allele name | Amplification from gDNA | Amplification from cDNA | Predicted DQB locus | Length of sequence (bp) |
|---|---|---|---|---|---|
| A | DQB2-like*03:01 | Yes (1) | Yes | DQB2-like | 894 |
| | DQB2*09:01 | Yes (1) | Yes | DQB2 | 897 |
| B | DQB1*02:01 | Yes (1) | NA | DQB1 | 887 |
| | DQB2*04:01 | Yes (1, 2) | NA | DQB2 | 891 |
| C | AJ238933 | Yes (1) | No | DQB1 | 270 |
| | LN868258 | Yes (2) | No | DQB2 | 270 |
| D[a] | Unknown A | Yes (1) | No | NA | NA |
| | Unknown B | Yes (2) | No | NA | NA |
| E | DQB2*10:01 | No | Yes | DQB2 | 870 |
| | DQB2-like*01:01 | Yes (1) | NA | DQB2-like | 802 |
| F | AJ23941 | Yes (1, 2) | No | DQB2 | 270 |
| | NewF | Yes (1, 2) | No | DQB1/DQB2[c] | 270 |
| G | DQB2*12:01 | No | Yes | DQB2 | 897 |
| | DQB2*09:01 | Yes (2) | Yes | DQB2 | 897 |
| | DQB2-like*01:01 | Yes (1) | NA | DQB2 | 802 |
| H[b] | NewH | Yes (1, 2) | No | DQB1 | 270 |
| | DQB2*11:01 | Yes (2) | Yes | DQB2 | 891 |

[a] No homozygous animals found so both gDNA and cDNA amplification attempts were from heterozygous animals

[b] Homozygous animals used for gDNA amplifications, but no RNA available for this haplotype thus cDNA amplification attempts made from heterozygous animals.

[c] Phylogenetic relationship unclear

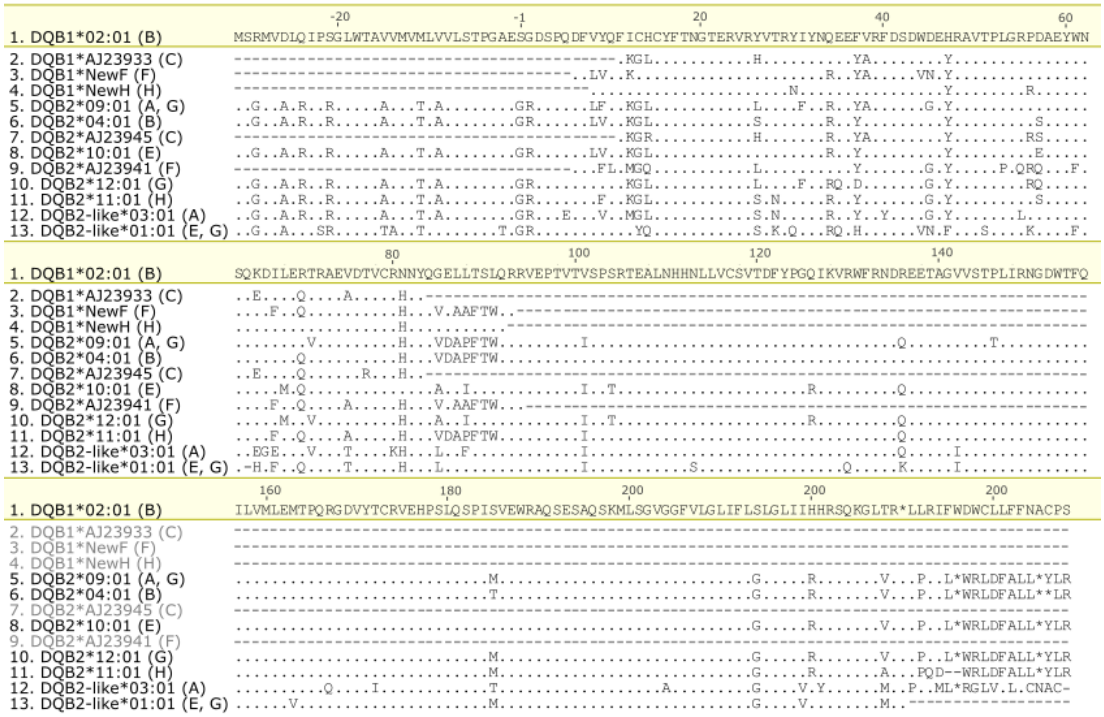


Figure 2.3. Predicted *DQβ* sequences identified in Soay sheep. Sequences are numbered from the first amino acid of the mature protein. Dots (.) indicate identity to sequence 1 and dashes (-) indicate missing sequence. The haplotypes on which each allele is present is shown in brackets after the allele name.

## 2.4.2 Phylogeny of *DQB* loci

Phylogenetic analysis of the *DQB* loci revealed three distinct clusters, corresponding to the three loci *DQB1*, *DQB2* and *DQB2-like* (Figure 2.4). All Soay *DQB* alleles are located within one of the three clusters. Alleles from each haplotype fell into two different clusters, with the following exceptions. Firstly, both of the haplotype F alleles fell into the *DQB2* cluster, although the allele NewF was not a full-length sequence and did pair with another *DQB1* sequence. Two alleles on haplotype G cluster within *DQB2*, however, this haplotype was found to have three alleles, and the third clustered most closely with *DQB2-like*.

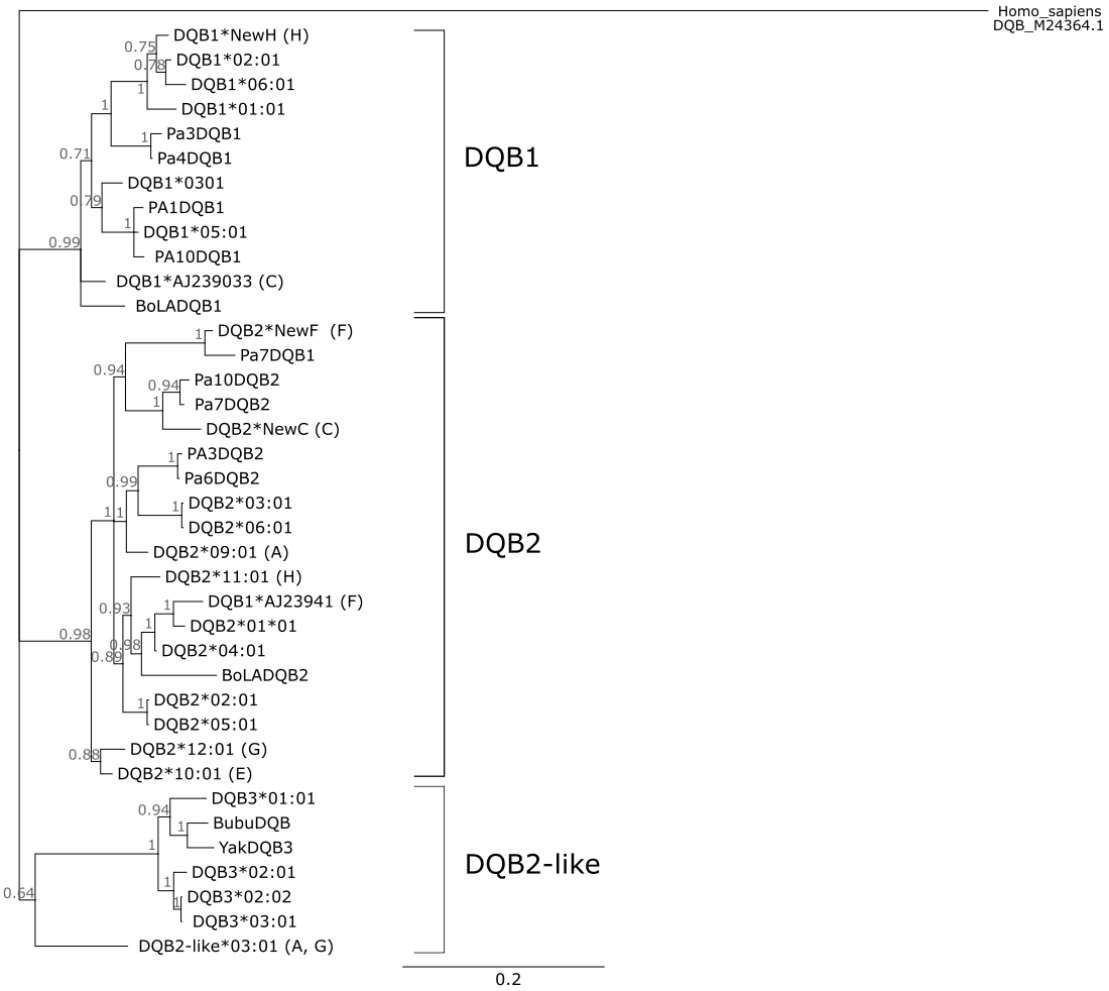



Figure 2.4. Phylogeny of *DQB* sequences estimated using MrBayes in Geneious. Letters in brackets indicate the Soay sheep haplotype that the allele was identified on. Numbers at branches are posterior probabilities. Alleles beginning Pa were not full-length *DQB* sequences (K. T. Ballingall, unpub).

### 2.4.3 Pairwise amino acid differences

The most similar haplotypes were haplotype G and either A or E each with 29 amino acid differences, depending on which of the two haplotype G alleles were included in the analysis (Table 2.4). This was also true when the *DQB* loci were excluded.

Table 2.4. Proportion of pairwise number of amino acid differences within exon 2 between haplotypes. Pairwise differences below the diagonal are for *DRB1*, *DQA* and *DQB*, of a total of 433 amino acids (haplotype D excluded due to lack of DQB sequences). Pairwise differences above the diagonal are for *DRB1* and *DQA*, of a total of 255 amino acids, to enable inclusion of haplotype D for which DQB alleles were not typed. Haplotype G was included in the *DQB* analysis twice, with either *DQB2*12:01* or *DQB2*09:01* as the *DQB2* allele.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| A | | 0.08 | 0.14 | 0.14 | 0.05 | 0.10 | 0.01 | 0.12 |
| B | 0.14 | | 0.11 | 0.08 | 0.10 | 0.05 | 0.06 | 0.07 |
| C | 0.21 | 0.16 | | 0.10 | 0.15 | 0.11 | 0.13 | 0.10 |
| D | NA | NA | NA | | 0.13 | 0.08 | 0.13 | 0.09 |
| E | 0.15 | 0.17 | 0.24 | | | 0.11 | 0.04 | 0.13 |
| F | 0.18 | 0.13 | 0.19 | NA | 0.21 | | 0.09 | 0.07 |
| G (*DQB2*12:01*) | 0.10 | 0.15 | 0.23 | NA | 0.07 | 0.21 | | 0.11 |
| G (*DQB2*09:01*) | 0.07 | 0.13 | 0.21 | NA | 0.08 | 0.18 | | |
| H | 0.19 | 0.10 | 0.16 | NA | 0.21 | 0.15 | 0.20 | |

### 2.4.4 Validation of sequences

Direct sequencing of *DRB1* and *DQA* loci in 94 individuals did not reveal any deviations from the allelic combinations identified using homozygous individuals. With the addition of the novel alleles on haplotypes C, F and H to the custom BLAST, alleles from all direct sequencing products could be determined, even from *DQA2*/*DQA2-like* heterozygous products from which four alleles co-amplified. Within these 94 individuals, 21.3 % were homozygous.

## 2.5 Discussion

Six *DRB1* alleles in Soay sheep were identified, which matched sequences previously identified in commercial Scottish sheep breeds and held in the IPD-MHC database. The level of homozygosity at the *DRB1* locus was 27.1 % of 118 individuals, which is higher than previously observed in other breeds of sheep (Stear *et al.* 2005; Herrmann-Hoesing *et al.* 2008a). However, at the class IIa haplotype level, this is an over estimate of homozygosity as some individuals homozygous for *DRB1*01:01* or *DRB1*22:01* are heterozygous at *DQA* and *DQB* loci. The level of class IIa haplotype homozygosity reduced to 21.3 % in the 94 validation samples genotyped across the *DRB1* and *DQ* loci. Note that these samples were originally selected to maximise representation of Soay

sheep diversity, so it is possible they are not totally representative of the population as a whole. No dramatic change in *DRB1* allele frequencies was observed over time, but samples numbers were too small to analyse temporal trends statistically. Further investigation at the population level will provide better estimates of the level of homozygosity and whether haplotype frequencies have changed over time within the Soay sheep population.

*DQA* and *DQB* genotyping revealed a total of eight class IIa haplotypes in the Soay sheep. *DRB1* and *DQA* genotyping from class IIa homozygous animals was relatively straight forward due to locus-specific primers (although the *DQA2* primer pair did co-amplify *DQA2-like* alleles). *DQB* primer pairs, however, showed varying levels of cross-amplification depending upon the alleles carried by the haplotype, which made phasing of alleles challenging without extensive cloning. The limited amount of class IIa haplotype diversity has made identifying variation at the *DQA* and *DQB* loci from homozygous animals possible. However, despite the limited variation, *DQB* allelic diversity from haplotype D could not be fully characterised due to the lack of available homozygous animals and limited numbers of heterozygous animals. Furthermore, assigning *DQB* alleles to loci was only possible using phylogenetic analysis. *DQB* genotyping would be improved by further development of locus-specific primers, although this is challenging due to limited genomic sequence in this region and the lack of locus-specific characteristics in exon 2 (van Oorschot *et al.* 1994; Wright & Ballingall 1994). Both known *DQ* haplotype configurations (*DQ1 + DQ2* and *DQ2 + DQ2-like*) were identified, as well as a novel configuration (two *DQB2* alleles with *DQA2 + DQA2-like + DQB2-like* on haplotype G) in Soay sheep. Both *DQB2* alleles on haplotype G were phylogenetically clustered within the *DQB2* group (Figure 2.4), and feature *DQB2* locus specific 3' UTR sequences. The two *DQB2* alleles were amplified from either cDNA (*DQB2*12:01*) or gDNA (*DQB2*09:01*). Whilst this might suggest that the *DQB2*09:01*allele is not transcribed, many of the *DQB* alleles identified from Soay sheep failed to amplify from cDNA. This probably reflects a combination of low cDNA quality and diversity in the primer binding sites. Whether the *DQB2*09:01*allele is transcribed or not remains uncertain. It cannot be ruled out that haplotype G carries three *DQB* loci. A previous study of class IIa haplotypes in Texel sheep identified a single haplotype with three *DQA* and three *DQB* loci and another with two *DQA* and three *DQB* loci (Ali *et al.* 2016). Variation in the number of *DQ* loci and *DQ* haplotype configurations may be more common than previously thought in *O. aries*.

An unusual allele was identified on haplotype A which was identical to the *DQB-E1* sequence described by Herrmann-Hoessing (Unpublished, Accession number HQ728697.1) and described here as *DQB2-like*03:01*. This sequence clustered with *DQB2-like* alleles, but did not carry the single codon deletion within the second exon described in other *DQB2-like* alleles (Ballingall *et al.* in press; Figure 2.3) but it does feature the typical *DQB2-like* 3' UTR sequence. As haplotype A carries the *DQA2 + DQA2-like* configuration, it would be expected that it also carries the *DQB2 + DQB2-like* configuration. Allele *DQB2-like*03:01* appears, therefore, be a divergent *DQB2-like* allele.

Numerous alleles were shared amongst the class IIa haplotypes within the Soay sheep. Pairwise alignment of the predicted amino acid sequences of second exons of the *DRB1*, *DQA* and *DQB* loci within each haplotype identified the greatest overall amino acid similarity between haplotypes A & G and A & E. However, the functional similarity of alleles is difficult to assess without detailed information on the diversity of peptides presented by each MHC molecule.

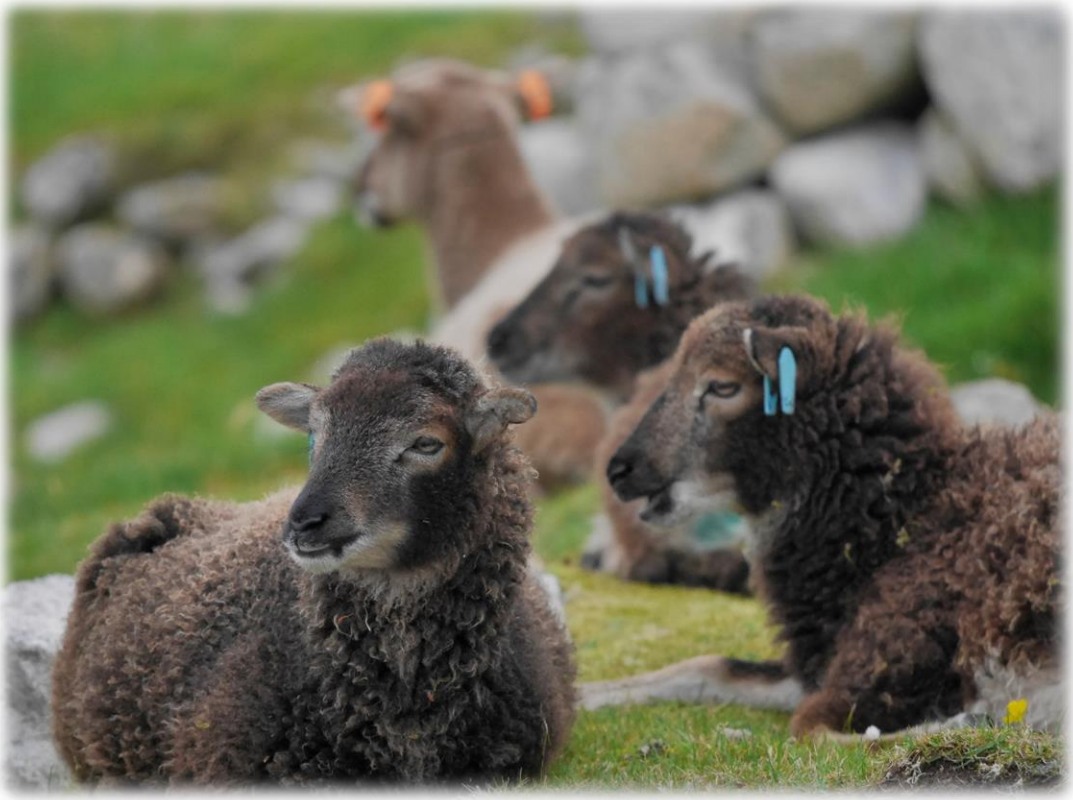
Variation in MHC class II molecules is formed from the combination of α and β chains. *DQ* molecules are therefore the products of polymorphic *DQA1 + DQB1* and *DQA2 + DQB2* genes (Ballingall *et al.* 2017 in press). *DQA2-like + DQB2-like* combinations are likely to provide additional functional diversity as this combination has been shown to express at the cell surface following co-transfection (Ballingall *et al.* 2017 in press). Haplotype E and G share the same *DQA2-like*01:01:01 + DQB2-like*01:01:01* allelic combination, so both haplotypes will generate the same *DQ2-like* molecule. Therefore, an individual homozygous or heterozygous for haplotypes E and G might have fewer *DQ* molecules compared to other haplotype combinations. On the other hand, haplotype G carries two *DQB2* genes. If both are capable of forming functional molecules in combination with *DQA2**01:02:01, it would have increased *DQ* molecule diversity. Intra-haplotype pairing of different *DQA* and *DQB* gene combinations may also provide additional class II molecules; however, not all allelic combinations are necessarily capable of generating functional molecules (Ballingall *et al.* 2017 in press).

The eight haplotypes characterised here are likely to be representative of all class IIa variation in the Soay sheep. The *DRA* locus, which shows only limited allelic diversity in *O. aries* (Ballingall *et al.* 2010), was not genotyped here. It is unlikely that genotyping the *DRA* locus would have further subdivided any haplotypes. Whilst the haplotypes from only a small number of primarily homozygous animals were characterised, the

haplotypes were validated at the *DRB1* and *DQA* loci in an additional 95 animals, and no new allelic combinations were identified. The *DQB* loci remain to be validated, and potential variation in the number of loci means that there is an unknown fraction of missing variation (Babik 2010). It is not certain that every haplotype has been detected, however, the extensive sequencing of multiple loci carried out throughout this study means that only alleles that occur at very low frequency will have been missed from our analysis.

Haplotype characterisation in the Soay sheep has revealed eight class IIa haplotypes of varying locus configurations. *DQA* and *DQB* loci were revealed to harbour substantial additional variation, and thus genotyping at just the *DRB1* locus would have disguised some of the class IIa variation. Similarly, the OLADRB microsatellite alone (Paterson 1998; Paterson *et al.* 1998) cannot detect the class IIa variation external to the *DRB1* locus (see Appendix C). Using the new in-depth knowledge of the class IIa haplotype diversity within the Soay sheep population, it may now be possible to develop a rapid SNP-based genotyping system in order to generate population level data. This would greatly facilitate analyses of evolutionary processes underlying the maintenance of the variation in the MHC class IIa region.

# 3. Selection and testing of a custom SNP panel for rapid determination of Class IIa haplotypes in large numbers of Soay sheep



## 3.1 Abstract

Investigating the evolutionary processes acting on a highly polymorphic gene region, such as the MHC, requires extensive population data on both phenotypes and genotypes. MHC variation exists as haplotypes of alleles at different loci. Due to the high level of polymorphism in the expressed genes of the MHC, it is notoriously difficult to devise genotyping methods targeting SNPs within these genes. Instead, in this chapter a fast genotyping system was developed, enabling imputation of class IIa haplotypes in Soay sheep using KASP genotyping of 11 SNPs located in the peripheral regions of the class IIa genes. These SNPs were identified from 188 Soays that were genotyped on the Ovine Infinium HD SNP Beadchip. The MHC class IIa haplotypes of 94 of these Soays were also determined using the genotyping-by-sequencing method described in Chapter 2. The Ovine Infinium HD SNP Beadchip also identified a new, recombinant haplotype. This recombination event had no effect on the functional class IIa haplotype but resulted in a new SNP haplotype which caused erroneous inference using only SNPs. To impute this new haplotype successfully, two SNPs from within exon 2 of the *DQA1* locus were added to the imputation panel. A total of 801 Soay sheep were successfully KASP genotyped, with 100% identity of genotypes returned for a control sample repeatedly genotyped. Of 37 individuals with haplotypes typed by Sanger sequencing and imputed from KASP SNPs, 36 gave identical haplotypes, and the incorrect haplotype inference was probably caused by a low genotyping rate. Out of the 801 individuals, 799 had diplotypes consisting of the (now) 9 known SNP haplotypes.

## 3.2 Introduction

Investigating the evolutionary processes acting on a highly polymorphic gene region, such as the MHC, requires extensive population data on both phenotypes and genotypes. Whilst gathering population level phenotypic data is a challenge in itself, collecting the corresponding high quality, locus-specific genotype data across multiple loci at the MHC in non-model animals is difficult, although becoming increasingly possible in mammals for class II genes (e.g. Bryja *et al.* 2007; Tollenaere *et al.* 2008; Li *et al.* 2014; Niskanen *et al.* 2014; Osborne *et al.* 2015). Locus-based genotyping methods are necessary to begin to disentangle the evolutionary processes maintaining the diversity of alleles across this region (Spurgin & Richardson 2010), in particular to

investigate the non-mutually exclusive mechanisms of heterozygous advantage, negative frequency-dependent selection and fluctuating selection. For many species, including many fish and birds, locus-based genotyping of the MHC is not possible due to gene duplication, gene conversion and variation in gene number make the design of locus-specific primers difficult (Westerdahl 2007; Babik 2010; Whittaker *et al.* 2012), but see (Worley *et al.* 2008)). This often results in studies co-amplifying loci and using pooled diversity measures (Westerdahl 2007; Worley *et al.* 2008; Babik 2010), but the inability to determine which locus an allele originated from prohibits assessment of the heterozygous state of a locus, and therefore the ability to effectively assess heterozygous advantage. On the other hand, genotyping only a single locus is quite unlikely to capture the full diversity of the MHC molecules generated by single haplotype (see Chapter 2). An effective and reliable genotyping method, therefore, should be able to detect the variation at individual loci and across multiple expressed loci.

Sequence-based genotyping methods have been developed for multiple ovine MHC class IIa loci (Ballingall & Tassi 2010; Ballingall *et al.* 2015, 2017 in press) which allow in-depth characterisation of variation in the region. The variation at class IIa loci was previously characterised in the Soay sheep population using these methods (Chapter 2). Genotyping of the *DRB1*, *DQA* and *DQB* loci revealed eight haplotypes, with some allele sharing amongst haplotypes. This means that no single locus captures the haplotypic variation in the Soay sheep population. In addition, sequence-based genotyping of the DQ loci, and in particular the DQB loci, was complicated by cross-amplification amongst loci and variation in locus configuration (Ballingall *et al.* in press, 2015; Ali *et al.* 2016; see Chapter 2). Furthermore, as methods, both Sanger and next-generation sequencing of multiple loci in large numbers of individuals would be costly and slow. Ideally, a method involving fast genotyping of key loci would be deployed to impute the haplotypes in each individual.

Human population studies have made use of SNP-based methods for imputing MHC haplotypes (Dilthey *et al.* 2011; de Bakker & Raychaudhuri 2012; Zheng *et al.* 2014) by carefully selecting a panel of SNPs which are in linkage disequilibrium with specific MHC haplotypes. High linkage disequilibrium between loci within subregions is characteristic of the mammalian MHC (Dawkins *et al.* 1999), including the ovine MHC (Lee *et al.* 2012). High levels of polymorphism and high sequence similarity between alleles at different loci within the class IIa genes (see Chapter 2) limits the ability to

select SNPs from within coding regions for which we have sequence data, and sequence data for introns and other non-coding regions of the ovine class IIa is generally lacking.

A total of 5805 Soay sheep have previously been genotyped on the Ovine SNP50 BeadChip (Illumina; 50K SNP chip) (Bérénos *et al.* 2014; Johnston *et al.* 2016), which includes approximately 37,000 SNPs which are variable in the Soay sheep. 188 key individuals within the Soay sheep pedigree, chosen for being representative of Soay diversity, have also been genotyped on the Ovine Infinium HD SNP BeadChip (HD SNP chip), which has an attempted 606,066 SNPs (Johnston *et al.* 2016). The MHC class IIa region itself was not expected to be well represented on either SNP chip, but it was hypothesised that there might be a sufficient number of SNPs in the flanking regions that were in linkage disequilibrium with the class IIa region to enable imputation of the Soay haplotypes. In this chapter, I therefore aimed to determine whether SNPs included on either the 50K SNP chip or the HD SNP chip were in linkage disequilibrium with the class IIa region, and if so, whether I could identify a minimal panel of SNPs that could be genotyped and used to impute the class IIa haplotypes.

## 3.3 Methods

### 3.3.1 Location of class IIa genes region with respect to SNPs on chips

To compare the SNP chip genotypes with class IIa haplotypes, it was first necessary to estimate where the class IIa genes were located on Chromosome 20. MHC genes are poorly mapped on the ovine genome due to the repetitive, multi-copy and polymorphic nature of MHC gene families. The *O. aries* genome (Oar_v3.1, GCA_000298735.1) on Ensembl (Yates *et al.* 2016) was used to search for protein families including the words "class II histocompatibility" in their description in order to find regions of the genome associated with classical class IIa loci and their homologues in other species. As the ovine MHC is known to be on chromosome 20, the results were limited to chromosome 20, as well as unmapped scaffolds. Additionally, the gene *BTNL2* was included as this is typically identified as the defining the end of the class IIa region (Gao *et al.* 2010; Liu *et al.* 2011; Lee *et al.* 2012)

### 3.3.2 50K and HD chip SNP locations in comparison to the putative class IIa gene region

To determine whether there are any SNPs within or around the class IIa gene region on the 50K SNP chip, PLINK v1.90 (Purcell *et al.* 2007) was used to select only SNPs on chromosome 20, to exclude SNPs which were fixed or had a call rate of less than 95 %, and to exclude individuals with less than 95 % of loci genotyped, leaving a total of 870 SNPs.

On the Ovine Infinium HD SNP BeadChip, there are 11,757 SNPs on chromosome 20. PLINK was used to perform quality control, excluding individuals with less than 95 % of loci genotyped, and SNP loci with minor allele frequencies (MAF) less than 1 % and call rates less than 95 %, leaving a total of 8420 SNPs.

### 3.3.3 Class IIa sequence-based genotyping of individuals typed by HD SNP chip

Sequence-based genotyping was previously carried out for the MHC class IIa loci (DRB1 and DQA) to determine class IIa haplotypes for 94 of the individuals which were included on the Ovine Infinium HD SNP Beadchip (Chapter 2). Here, sequence-based genotyping at the DRB1 and DQA1 loci using the same methodology as Chapter 2 was carried out for the remaining 94 individuals that had been genotyped on the HD SNP chip. Without genotyping the DQA2 and DQA2-like loci, it was not possible to distinguish between haplotypes A and G.

### 3.3.4 Identification of the class IIa linkage region

Using HD SNP genotypes for the 188 individuals, pairwise linkage disequilibrium (LD) between the HD SNPs (within 500 kb proximity) was calculated in Haploview (Barrett *et al.* 2005) as both $r^2$ and $D'$ between positions 24.5 and 26.5 Mb on chromosome 20, to encompass the 25.353 – 25.784 Mb region identified as containing the class IIa genes (see results). LD blocks were identified using the Four Gamete Rule (Wang *et al.* 2002). The Four Gamete Rule estimates the population frequencies of the four possible combinations for a pair of SNPs (the four gametes), and recombination is deemed to have occurred if all four frequencies are greater than 0.01; a block is defined from consecutive markers where only three of the possible four gametes are observed in the population (Barrett *et al.* 2005). Blocks were numbered according to their proximity to the HD SNP gap in the centre of the class II region, with those upstream (i.e. closer to

the centromere) of this gap being decreasingly negative and those downstream (i.e. closer to the q-arm telomere) being increasingly positive.

Next, the HD SNP haplotypes were compared to the class IIa Sanger sequencing haplotypes. The seven HD SNPs within block –1 as identified by the Four Gamete Rule (immediately upstream of the class IIa gap) were phased using BEAGLE (Browning & Browning 2007) with standard settings. The phased haplotypes were compared with Sanger class IIa haplotypes. Upstream HD SNPs were then added sequentially and cumulatively to block –1 haplotypes, and phased using BEAGLE after each SNP addition. This identified if and where the HD SNP and Sanger haplotypes matched, and where the LD degraded, i.e. that a single Sanger haplotype was represented by multiple SNP haplotypes. The process was repeated for downstream HD SNPs. The class IIa linkage region was identified as a continuous region of HD SNPs, spanning the SNP gap, for which the phased SNP haplotypes matched the Sanger haplotypes.

### 3.3.5 Haplotype B and H recombinant detection

Whilst determining the class IIa linkage region, a recombination between haplotypes B and H was identified (see Results). The recombinant individual was haplotype B across the class IIa loci (determined by Sanger sequencing) but its SNP haplotype could not be distinguished from haplotype H using the SNPs within the class IIa linkage region. Therefore, candidate intragenic SNPs were identified from the sequences of exon 2 of the *DRB1* and *DQA* loci that were capable of separating Sanger haplotypes B and H. *DQB* genes were not included due to the lack of available sequence data for haplotype D. Candidate intragenic SNPs were identified from allelic sequences generated from Soay sheep in Chapter 2, were bi-allelic across all loci (when the locus was present on a haplotype) and required a minimum of 50 bp of flanking sequence on either side of the SNP. Whilst candidate SNPs were required to be bi-allelic, *DQA1* SNPs are always truly tri-allelic, where the third allele, a null allele, can be considered as the absence of the SNP on haplotypes where the *DQA1* locus is not present.

### 3.3.6 KASP SNP panel selection

Kompetitive allele specific PCR (KASP) system (LGC Genomics, Hoddesdon) can be used to assay individual SNPs. Candidate SNPs, including both the HD SNPs and intragenic SNPs, were sent for *in-silico* primer validation using the Kraken software by LGC. Flanking sequences of the selected HD SNPs were downloaded from dbSNP

(http://www.ncbi.nlm.nih.gov/SNP/). Flanking sequences of intragenic SNPs within the class IIa loci were obtained from the sequences generated in Chapter 2, with polymorphic sites identified only from alleles known from the Soay sheep population.

All HD SNPs and only two intragenic SNPS (both within *DQA1*) were approved by *in-silico* primer validation. HD SNP alleles and allelic combinations were identified which were unique to each class IIa Sanger haplotype, preferentially selecting those closest to the SNP gap in the centre of the class IIa region. We then identified a minimum subset of HD SNPs which were able to impute the two class IIa haplotypes (diplotype) of an individual whilst maintaining degeneracy. For each haplotype, there were two HD SNPs that distinguished the class IIa haplotype, and thus it should still be possible to determine haplotypes if the alternative HD SNP failed to genotype. SNP oar3_OAR20_27259292.1 was not included as it is already present on the Ovine50K SNP BeadChip and has therefore already been genotyped in the Soay sheep population (Bérénos *et al.* 2015), but was not used in imputation of haplotypes. Both *DQA1* intragenic SNPs were included on the final genotyping panel.

### 3.3.7 KASP genotyping

The final panel of SNPs used for KASP genotyping included 11 HD SNPs and two intragenic *DQA1* SNPs, as well as an additional 11 non-MHC SNPs selected by other project members for non-MHC studies. This panel was genotyped by LGC Genomics in 941 animals born between 1998 and 2007. DNA was extracted previously for genotyping on the Ovine50K SNP BeadChip using Qiagen DNeasy kit in 96-well plate format (Johnston *et al.* 2011; Bérénos *et al.* 2014). One individual was included on every plate in a random well to confirm genotyping consistency, and one No Template Control (NTC) was included on each plate. Samples were submitted to LGC for KASP genotyping, and calling was performed automatically by LGC following standard protocol and the data returned. The 11 non-MHC SNPs were removed from the data set before subsequent analysis.

PLINK v1.90 (Chang *et al.* 2015) was used to exclude individuals with more than 50% missing genotypes across the 11 HD SNPs, and then Beagle v4.0 (Browning & Browning 2007) was used to phase the genotypes. The phased HD SNP haplotypes were then matched to the Sanger haplotypes. Because the *DQA1* intragenic SNPs were tri-allelic (major allele, minor allele and a null allele), they were not included in the phasing process. On the KASP-genotyped plates there were 37 individuals that had also been

previously genotyped by the Ovine Illumina HD BeadChip, and the phased haplotypes were compared to ensure consistency of the methods.

### 3.3.8 Detection of recombinant individuals using intragenic DQA1 SNPs

*DQA1* SNP genotypes were compared to the SNP haplotypes to identify individuals carrying conflicting profiles, that is, the phased HD SNP profile of haplotype H but the DQA1 genotype of haplotype B. This enabled the identification of B to H recombinant individuals. To assess whether the *DQA1* SNP genotypes were reliable, expected *DQA1* genotype profiles were compared to SNP haplotypes for all diplotype combinations. For *DQA1*null* haplotypes, the *DQA1* locus is absent and therefore there is the potential for primers to cross-amplify with other regions of the genome, including pseudogenes and other loci such as *DQA2* and *DQA2-like*, with highly similar nucleotide sequences.

## 3.4 Results

### 3.4.1 Genomic location of class IIa gene region

Nine protein families represented by 22 transcripts were identified by searching the Oar_v3.1 genome sequence on the Ensembl database with the term "class II histocompatibility" in their descriptions, as well as *BTNL2* (Table 3.1). Five transcripts were located on unmapped scaffolds. The remaining 17 transcripts were clustered in two regions of chromosome 20, 7.164 – 7.426 Mb and 25.353 – 25.812 Mb, with the former including primarily class IIb gene annotations (DO and DM genes) and the latter class IIa gene annotations (DR and DQ genes). DQA gene annotations were found in both class II regions, which likely reflects poor sequence assembly or gene annotation in this highly repetitive and polymorphic region. Previous assemblies of this region using BAC cloning did not identify DQ loci outside of the class IIa region (Liu *et al.* 2006; Gao *et al.* 2010). The classical class IIa genes are therefore expected to be located between 25.353 and 25.812 Mb.

### 3.4.2 50K and HD SNP genotyping

On the 50K SNP chip, there were 1053 SNPs distributed across chromosome 20, of which 870 passed quality control. There were only ten SNPs excluded due to low genotyping rates (< 95 % of individuals), and 173 SNPs excluded because they were

fixed in all individuals. Only a single SNP occurred within the class IIa linkage region (Figure 3.1a).

On the HD SNP chip, there were 11,757 HD SNPs located on chromosome 20, of which 8420 SNPs passed quality control. Of these, 21 were located within the class IIa linkage region (Figure 3.1b), and no SNPs were excluded within this region because of low MAF. There are no SNPs between 25.46 MB and 25.685 Mb in the centre of the class IIa gene region, which is known from here on as the HD SNP gap. All 188 sheep passed quality control (> 95 % of SNPs genotyped per individual).

Table 3.1. Locations of transcripts within protein families containing “class II histocompatibility” in the description occurring on chromosome 20 of the *O. aries* genome (Oar_v3.1, GCA_000298735.1) or on unmapped scaffolds.

| Protein Family ID | Description | Chromosome/ scaffold | Min position (bp) | Max position (bp) | Additional annotation | Protein ID |
|---|---|---|---|---|---|---|
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | AMGL01121385.1 | 21 | 5,321 | | ENSOARP00000017463 |
| PTHR19944_SF52 | Class II histocompatibility antigen beta chain precursor | JH922951.1 | 480 | 710 | Human homologue DQB | ENSOARP00000000037 |
| PTHR19944_SF54 | HLA Class II histocompatibility antigen DQ alpha chain precursor | AMGL01119849.1 | 1,997 | 5,240 | DQ | ENSOARP00000003175 |
| PTHR19944_SF52 | Class II histocompatibility antigen beta chain precursor | JH922521.1 | 2,724 | 3,046 | Human homologue DQB | ENSOARP00000020450 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | JH923254.1 | 3,431 | 4,459 | | ENSOARP00000002509 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | 20 | 7,164,298 | 7,178,949 | | ENSOARP00000007745 |
| PTHR19944_SF55 | HLA class II histocompatibility antigen, DQ alpha 2 chain precursor | 20 | 7,198,200 | 7,208,505 | DQA | ENSOARP00000007864 |
| PTHR19944_SF55 | HLA class II histocompatibility antigen, DQ alpha 2 chain precursor | 20 | 7,198,200 | 7,204,636 | DQA | ENSOARP00000007865 |
| PTHR19944_SF52 | Class II histocompatibility antigen beta chain precursor | 20 | 7,221,436 | 7,228,230 | Human homologue DQB | ENSOARP00000007912 |
| PTHR19944_SF43 | Class II histocompatibility antigen DO beta chain precursor MHC class II antigen DOB | 20 | 7,234,829 | 7,248,527 | DOB | ENSOARP00000008004 |
| PTHR19944_SF51 | Class II histocompatibility antigen beta chain precursor | 20 | 7,346,814 | 7,357,333 | DMB | ENSOARP00000008482 |
| PTHR19944_SF13 | Class II histocompatibility antigen alpha chain precursor | 20 | 7,367,386 | 7,372,035 | DMA | ENSOARP00000008555 |
| PTHR19944_SF44 | HLA class II histocompatibility antigen DO alpha chain precursor MHC DN alpha MHC DZ alpha MHC class II antigen DOA | 20 | 7,420,413 | 7,425,930 | DOA | ENSOARP00000008678 |
| PTHR19944_SF54 | HLA Class II histocompatibility antigen DQ alpha chain precursor | 20 | 25,353,186 | 25,356,656 | DQ | ENSOARP00000016611 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | 20 | 25,398,827 | 25,402,957 | | ENSOARP00000016782 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | 20 | 25,450,028 | 25,460,293 | | ENSOARP00000016804 |
| PTHR19944_SF55 | HLA class II histocompatibility antigen, DQ alpha 2 chain precursor | 20 | 25,501,894 | 25,572,798 | DQA | ENSOARP00000016862, ENSOARP00000016864 |
| PTHR19944_SF52 | Class II histocompatibility antigen beta chain precursor | 20 | 25,532,506 | 25,543,103 | Human homologue DQB | ENSOARP00000017039 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | 20 | 25,594,470 | 25,608,591 | DRB3 | ENSOARP00000017287 |
| PTHR19944_SF55 | HLA class II histocompatibility antigen, DQ alpha 2 chain precursor | 20 | 25,666,699 | 25,676,792 | DQA | ENSOARP00000017463 |
| PTHR19944_SF45 | H2 Class II histocompatibility antigen E beta chain precursor | 20 | 25,736,126 | 25,741,866 | | ENSOARP00000017715 |
| PTHR19944_SF47 | Class II histocompatibility antigen DR alpha chain precursor MHC class II antigen DRA | 20 | 25,774,090 | 25,783,626 | DRA | ENSOARP00000017840 |
| PTHR24100_SF5 | Butyrophilin like 2 | 20 | 25,793,691 | 25,811,932 | BTNL2 | ENSOARG00000016780 |

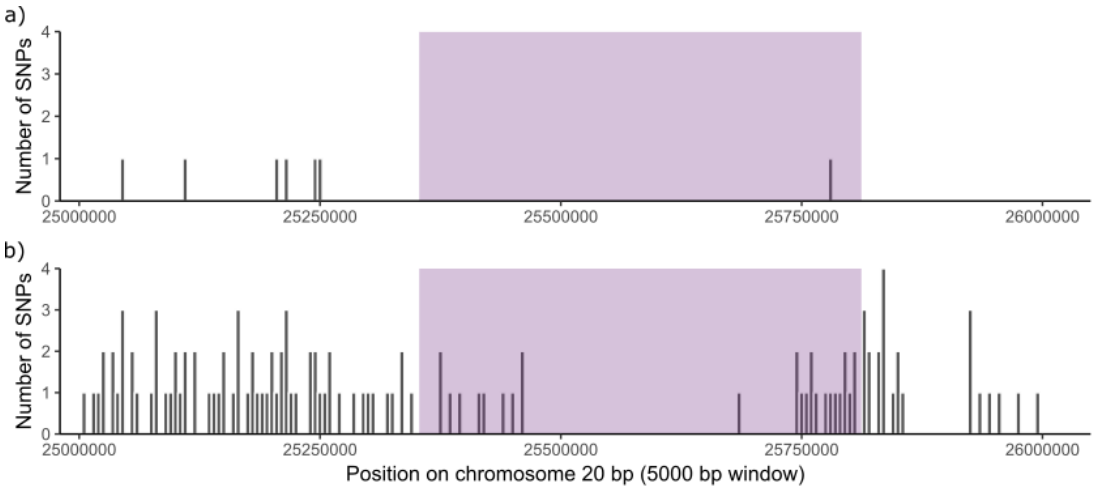



Figure 3.1. Number of SNPs within each 5000 bp window on chromosome 20 between 25 and 26 Mb on a) the 50K SNP chip and b) the HD SNP chip. The purple block represents the class IIa gene region between 25.353 to 25.812 Mb, as defined above.

### 3.4.3 Class IIa sequence-based genotyping

Full class IIa diplotypes were determined by sequence-based genotyping for 135 out of the 188 individuals genotyped on the Ovine Illumina HD BeadChip (this includes the 94 individuals genotyped in Chapter 2). Partial diplotype information was determined for an additional 50 individuals which carried haplotypes A or G, but these haplotypes were not differentiated. Three individuals failed sequence-based genotyping.

### 3.4.4 Identification of the class IIa linkage region

Using all the HD SNP chip genotypes in the estimated class IIa gene region for all 188 sheep, pairwise LD estimates of $D'$ were typically very high, although not in all cases (Figure 3.2). There was no obvious pattern to the degradation of LD, which would be suggestive of a recombination hotspot. $R^2$ estimates were typically very low (Figure 3.3), compared to $D'$ estimates, and were also uninformative in identifying linkage regions.

Given the lack of obvious LD pattern, linkage blocks identified using the Four Gamete Rule in Haploview (shown in Figure 3.2 and Figure 3.3) were used to identify a single continuous region which corresponded to each of the eight class IIa haplotypes. Block –1, which is immediately upstream of the HD SNP gap, formed six haplotypes (Figure 3.4). Four of the six haplotypes in block –1 each corresponded to a single Sanger haplotype, and two each corresponded to a pair of haplotypes, A and G or B and H

(Figure 3.4Figure 3.2). The addition of upstream SNPs to block –1 showed that linkage degraded immediately upstream of block –1 (at SNP oar3_OAR20_25383037) within haplotype E, but haplotypes A and G, as well as haplotypes B and H, only became differentiated at oar3_OAR20_25346061 (block –3). Downstream of the SNP gap, haplotypes B and H became differentiated at oar3_OAR20_25742763 (block 1), and A and G differentiated with the addition of oar3_OAR20_25752381 (adjacent to block 1). Linkage within haplotypes did not degrade until oar3_OAR20_25833644 (within block 6) for both haplotypes B and E (Figure 3.4). Thus, a continuous linkage region of 31 HD SNPs was identified between oar3_OAR20_25393230 and oar3_OAR20_25831572, which encompasses 438.34 kb.

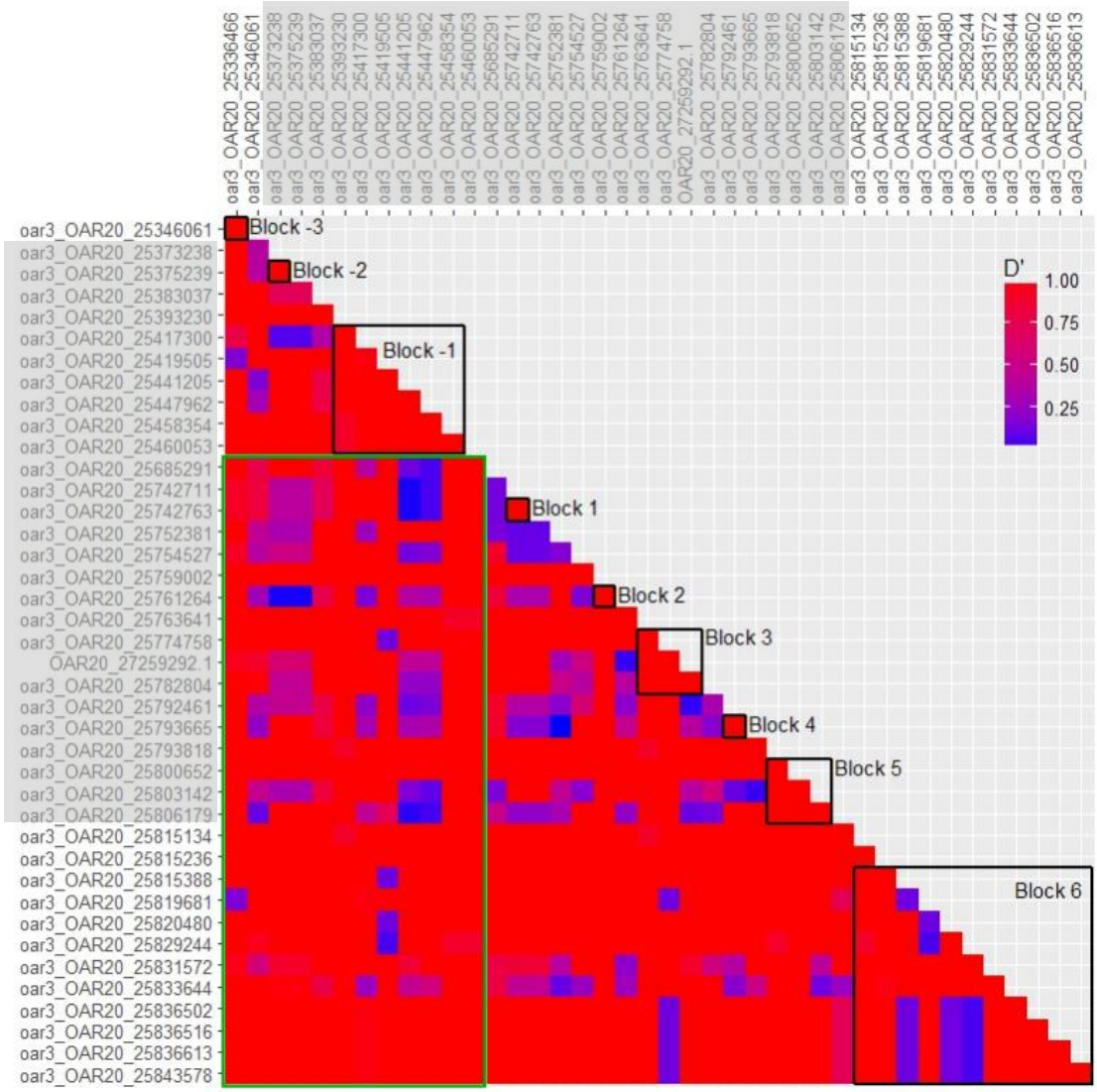



Figure 3.2. *D'* linkage disequilibrium estimates for pairs of HD SNPs within and surrounding the class IIa gene region, where values of 1 represents high LD, and 0 represents low LD. SNPs within the class IIa gene region are highlighted by the grey shading. The HD SNP gap is denoted by the green box, and SNP pairs that fall within this green box represent pairs spanning the SNP gap (i.e. that one SNP is upstream and the other downstream). Open black boxes show linkage blocks identified within Haploview using the Four Gamete Rule.

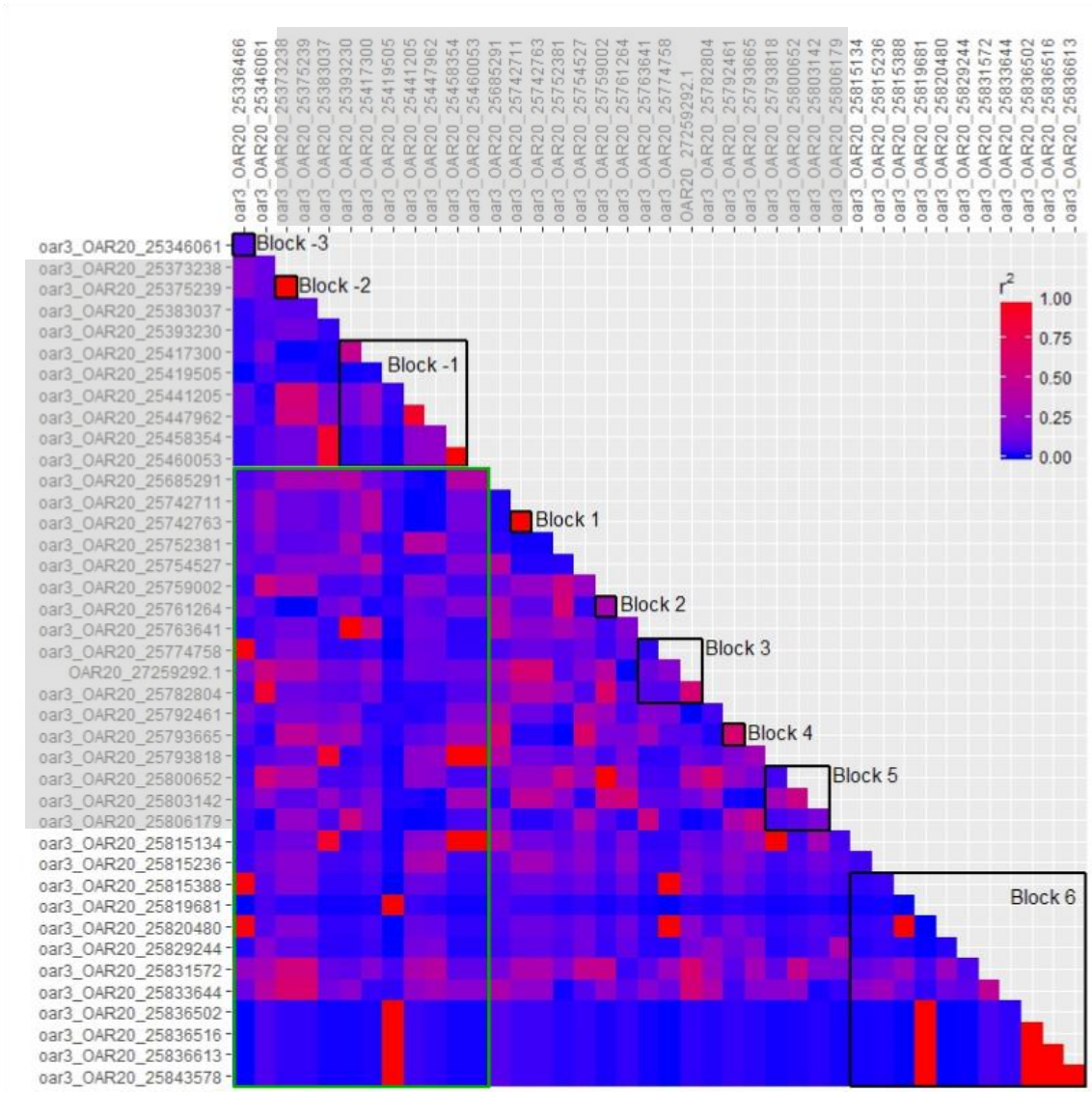



Figure 3.3. $r^2$ linkage disequilibrium estimates for pairs of HD SNPs within and surrounding the class IIa gene region, where values of 1 represents high LD, and 0 represents low LD. SNPs within the class IIa gene region are highlighted by the grey shading. The HD SNP gap is denoted by the green box, and SNP pairs that fall within this green box represent pairs spanning the SNP gap (i.e. that one SNP is upstream and the other downstream). Open black boxes show linkage blocks identified within Haploview using the Four Gamete Rule.

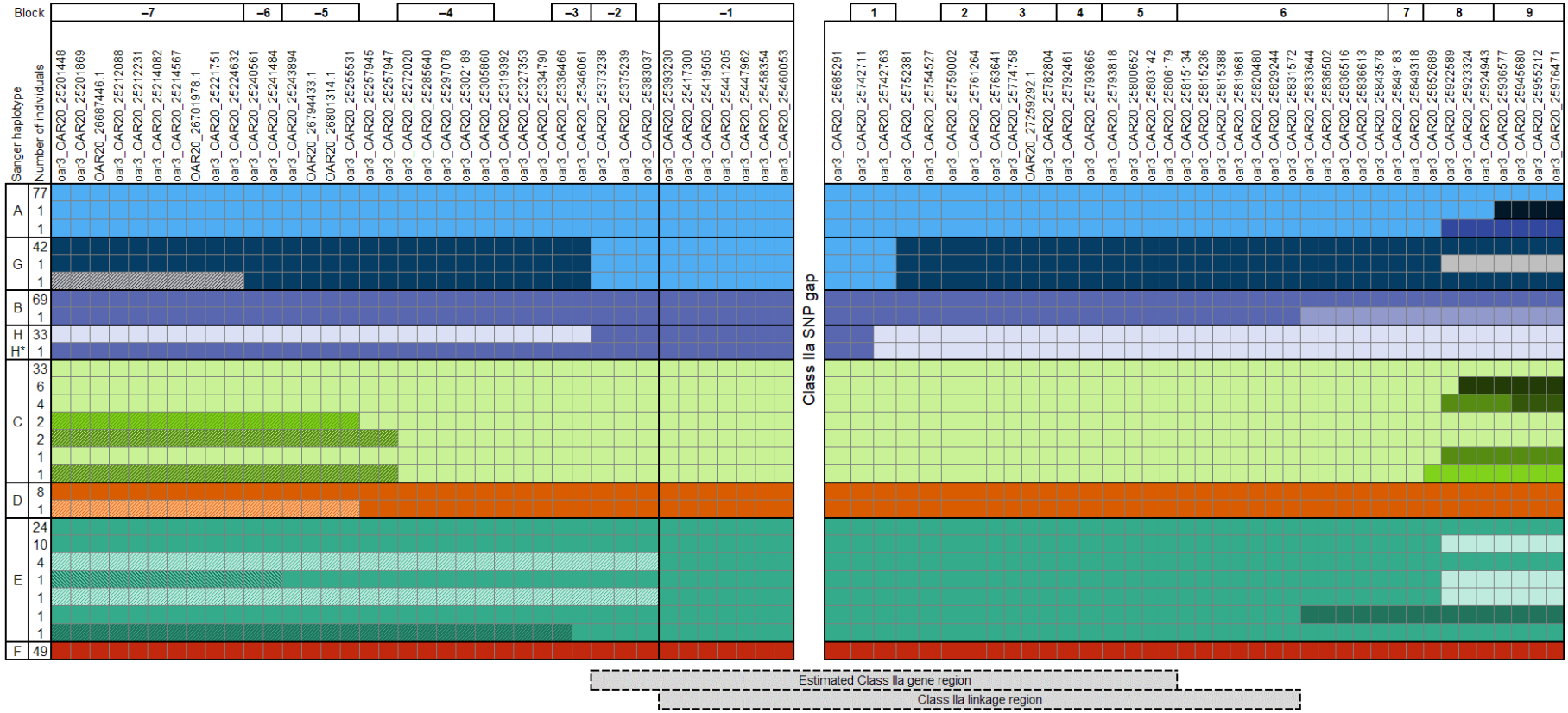



Figure 3.4. Representation of the phased SNP haplotypes (rows) formed by sequentially and cumulatively adding either upstream (left) or downstream (right) HD SNPs to block –1. Linkage blocks (Four Gamete Rule in Haploview) are indicated above the SNP names, and corresponding Sanger haplotypes in the left-most column. Moving outwards from block –1, a colour change along the row indicates that an additional haplotype is formed with the addition of that SNP. The estimated class IIa gene region (25.353 to 25.812 Mb) and the class IIa linkage region are indicated below the plot. H* is a recombinant haplotype detected in a single individual, for which the SNP haplotype within the class IIa linkage region matches haplotype H but class IIa Sanger sequencing is haplotype B.

### 3.4.5 Haplotype B and H recombinant detection

Individual 4179 was found to be heterozygous across Sanger haplotypes for B and G. The HD SNPs, however, revealed the existence of a recombination event which did not affect the Sanger class IIa haplotype, but affected the haplotype that would be imputed using the panel of 31 HD SNPs within the class IIa linkage region. Using only the SNPs upstream of the SNP gap, 4179's SNP haplotype was consistent with haplotypes B and G (Figure 3.4, haplotype H*). Using only the SNPs downstream of the SNP gap, 4179's SNP haplotype was consistent with haplotypes H and G. Therefore, using the SNPs within the class IIa linkage region (i.e. block –1 and downstream SNPs), the haplotype would be incorrectly imputed as H. This indicates the existence of a recombination event between haplotypes B and H which has occurred downstream of the class IIa genes included in the Sanger haplotype (the *DRB1, DQA* and *DQB* genes).

The recombination event cannot be detected using SNPs within the class IIa linkage region (see Figure 3.5). Potential SNPs within class IIa genes that were capable of distinguishing haplotypes B and H were therefore identified from the allelic sequences generated in Chapter 2. Four *DRB1* SNPs, 12 *DQA1* SNPs, and ten *DQA2/2-like* SNPs were identified (Figure 3.6).

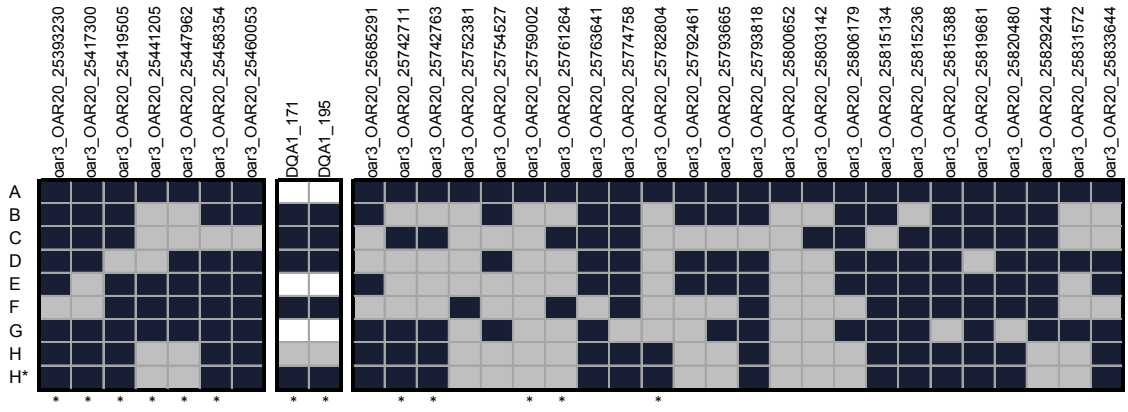



Figure 3.5. SNP alleles (columns) within the class IIa linkage region for each Sanger haplotype (rows). Dark cells indicate that the allele at that SNP was identical to haplotype A, and grey cells indicate the allele differs. Genotypes for the *DQA1* SNPs which were approved by KASP *in-silico* validation (see Figure 3.6) are shown in the position of the HD SNP gap. White cells correspond to *DQA1*null* alleles on haplotypes where the D*QA1* locus is absent. *DQA1*null* alleles are therefore the absence of amplification. Stars beneath the SNP column indicate those SNPs selected for the KASP assay. SNP loci are shown in order of their position on chromosome 20.

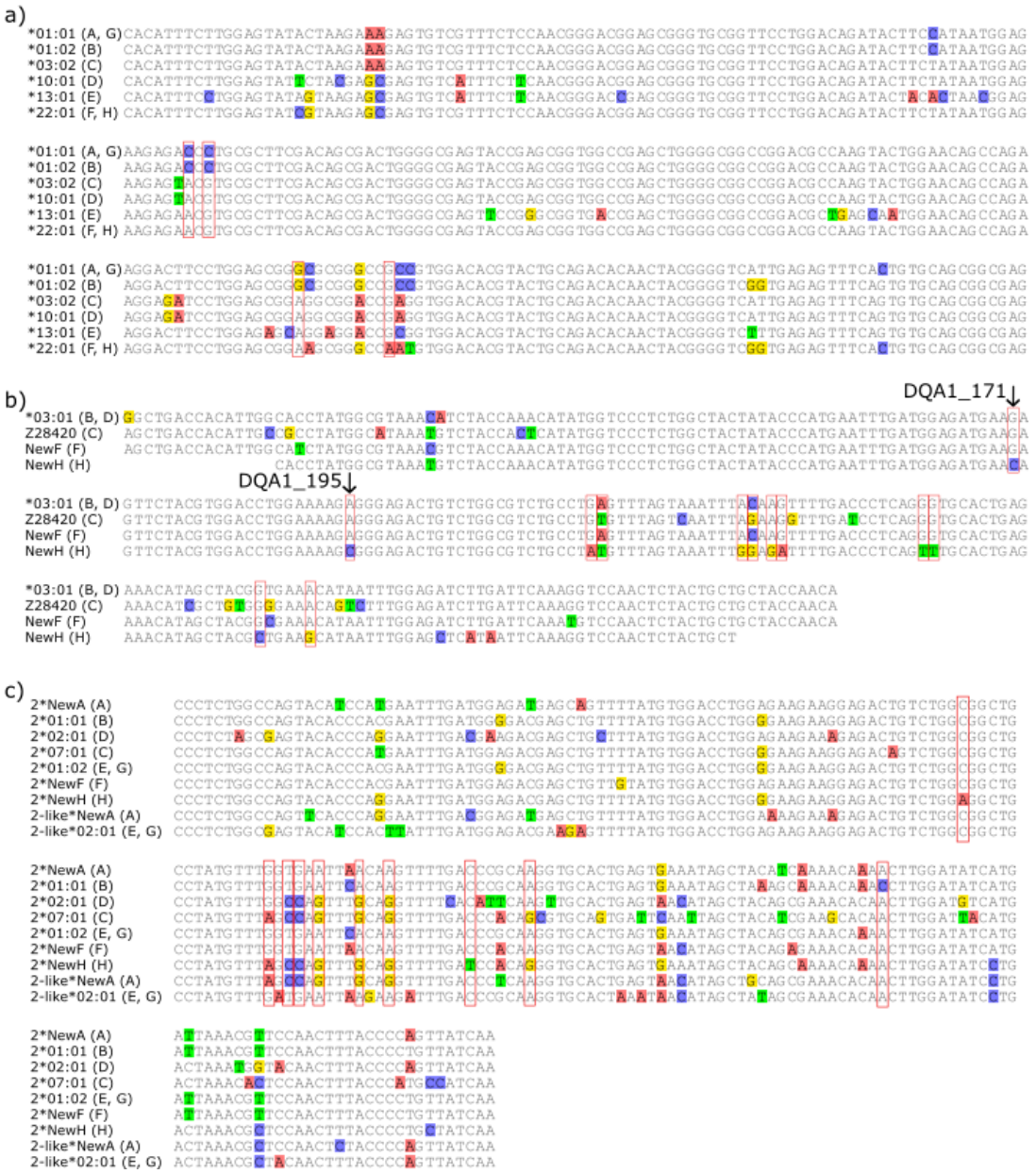



Figure 3.6. Aligned nucleotide sequences for exon 2 of a) DRB1, b) DQA1 and c) DQA2 and DQA2-like alleles within the Soay sheep population. SNPs which distinguish haplotypes B and H are highlighted by red boxes. SNPs were only considered if they were bi-allelic across all alleles and at least 50 bp of flanking sequence on either side of the SNP was known. SNPs approved by KASP *in-silico* are indicated by the arrows and SNP ID is shown beside.

### 3.4.6 KASP assay selection, genotyping and haplotyping

All 31 HD SNPs within the class IIa linkage region that were selected from the Ovine Infinium HD SNP BeadChip passed *in silico* validation for KASP. Of the intragenic SNPs from the Sanger sequencing designed to detect the recombinant haplotype, only two (DQA1_171 and DQA1_195) were approved, and these were the only SNPs submitted for this purpose with invariant flanking sequences (Figure 3.6). From this shortlist, we were able to impute the eight class II haplotypes using 11 HD SNPs and detect the haplotype B to H recombination, whilst maintaining degeneracy in case of genotyping failures.

Quality control was carried out using only the 11 HD SNPs, and 801 out of 941 individuals were retained after excluding individuals with more than 50% missing genotypes. Genotype calls for the individual repeated on each DNA plate were identical for all loci, thus confirming repeatability of the assays. MAFs for the 11 HD SNPs range from 0.116 (oar3_OAR20_25458354) to 0.493 (oar3_OAR20_25742711), all loci had a genotyping success rate of greater than 95% of individuals and no HD SNPs were found to deviate from HWE following Bonferroni correction (Table 3.2).

Phasing of the 11 KASP genotyped HD SNPs in the 801 individuals revealed 10 haplotypes (Table 3.3), eight of which matched the expected class IIa haplotypes and two that were novel. The two novel haplotypes were each identified in a single individual, which were missing zero and one genotypes. Of the 37 individuals genotyped by both the Ovine Illumina HD BeadChip and KASP methodologies, the phased haplotypes for 36 individuals were identical. For one individual, 4399, one of the two haplotypes was imputed incorrectly from the KASP genotypes (true haplotype B imputed incorrectly as haplotype E), although genotyping failed at four loci in that individual.

Table 3.2. Frequencies of phased haplotypes using the panel of 11 HD SNPs.

| SNP haplotype | Count | Sanger haplotype | Novel SNP haplotype | Count |
|---|---|---|---|---|
| AACCGGATGCG | 297 | A | AGCTGGGCAAA | 1 |
| AACTAGGCAAA | 344 | B | GACTAGGCAAA | 1 |
| AACTACATACA | 181 | C | | |
| AATTGGGCAAA | 55 | D | | |
| AGCCGGGCAAA | 152 | E | | |
| GGCCGGGCACA | 230 | F | | |
| AACCGGATAAA | 192 | G | | |
| AACTAGATAAG | 149 | H | | |

Table 3.3. Locus information, genotyping success rate (%), MAF (minor allele frequency), $H_O$ (observed heterozygosity), $H_E$ (expected heterozygosity), and Hardy-Weinberg Equilibrium (HWE) significance values, for the SNPs selected from the Ovine Infinium HD BeadChip ID that were genotyped by KASP.

| SNP ID | rs accession number[b] | Genome Position | Major Allele | Minor Allele | Genotyping success rate (%) | MAF | $H_O$ | $H_E$ | HWE p value |
|---|---|---|---|---|---|---|---|---|---|
| oar3_OAR20_25759002 | rs401547149 | 25759002 | G | A | 98.38 | 0.170 | 0.286 | 0.282 | 0.802 |
| oar3_OAR20_25458354 | rs403182119 | 25458354 | C | G | 99.38 | 0.123 | 0.205 | 0.215 | 0.327 |
| oar3_OAR20_25447962 | rs407228358 | 25447962 | A | G | 97.63 | 0.439 | 0.499 | 0.492 | 0.809 |
| oar3_OAR20_25393230 | rs408059130 | 25393230 | G | A | 99.25 | 0.145 | 0.254 | 0.248 | 0.636 |
| oar3_OAR20_25742711 | rs408458106 | 25742711 | G | A | 97.25 | 0.485 | 0.490 | 0.500 | 0.670 |
| oar3_OAR20_25441205 | rs412204478 | 25441205 | T | C | 98.38 | 0.473 | 0.517 | 0.499 | 0.476 |
| oar3_OAR20_25417300 | rs412489249 | 25417300 | G | A | 97.88 | 0.243 | 0.335 | 0.368 | 0.060 |
| oar3_OAR20_25761264 | rs414259673 | 25761264 | C | A | 98.25 | 0.442 | 0.502 | 0.493 | 0.738 |
| oar3_OAR20_25419505 | rs419818757 | 25419505 | T | C | 98.38 | 0.031 | 0.063 | 0.061 | 0.677 |
| oar3_OAR20_25742763 | rs421513037 | 25742763 | C | T | 97.75 | 0.490 | 0.494 | 0.500 | 0.813 |
| oar3_OAR20_25782804 | rs422511469 | 25782804 | G | A | 94.51 | 0.279 | 0.403 | 0.402 | 0.952 |
| DQA1_171[a] | NA | 25600000 | C | G | 83.52 | 0.146 | 0.118 | 0.249 | <0.0001 |
| DQA1_195[a] | NA | 25600300 | C | A | 82.77 | 0.148 | 0.116 | 0.252 | <0.0001 |

[a] Includes *DQA1**null allele which is not accounted for in genotyping success rate, $H_O$, $H_E$, HWE calculations or MAF
[b] Accession number in Ensembl http://www.ensembl.org/Ovis_aries/

### 3.4.7 DQA1 SNP genotypes

An individual's KASP genotype for a particular SNP is called based on the fluorescence cluster it is falls within, and a typical example of this is shown in Figure 3.7a. A homozygote emits only FAM or HEX fluorescence according to the allele it carries, and heterozygotes emit approximately half FAM and half HEX. *DQA1* SNPs were more complex in that there are three potential alleles: major allele, minor allele and null allele. *DQA1*null* is the absence of the *DQA1* locus on a haplotype, and the SNP assay fails to amplify. *DQA1*null* homozygotes would be expected to cluster with the No Template Controls (NTCs) and samples that failed to amplify for other reasons, and *DQA1*null* heterozygotes to cluster with the homozygotes of the alternative haplotype.

Despite the presence of null alleles at the *DQA1* loci, clustering remained efficient (Figure 3.7b and Figure 3.7c), although the scatter amongst the NTCs was more pronounced than for a typical biallelic SNP (Figure 3.7a). There were 11 recombinant individuals identified by comparison of the expected *DQA1* genotypes according to the imputed SNP haplotype with the actual *DQA1* genotypes (Figure 3.8). The expected and actual *DQA1* genotypes were in agreement for all 37 individuals genotyped by both Ovine Infinium HD BeadChip and KASP methods, including generating an accurate *DQA1* genotype for individual 4399, whose diplotype was incorrectly imputed by the KASP method.

Five individuals had unexpected genotypes (Figure 3.8) – that is, their actual DQA1 genotype was not as expected according to their diplotype (Table 3.4). Four individuals were imputed as carrying two haplotypes which are *DQA1*null* and therefore *DQA1* SNP loci should have failed to amplify; however, amplification was detected at one or both *DQA1* SNP loci. One individual carried haplotypes A (*DQA1*null*) and D, so the expected *DQA1* SNP genotype was as for haplotype D homozygotes (G:G or A:A for DQA1_171 and DQA_195 respectively); however, the actual *DQA1* SNP genotypes were as expected for haplotype H homozygotes (C:C for both DQA1 loci). Only one of these five individuals had a SNP genotype imputed using less than the full panel of 11 HD SNPs (4 SNPs failed). *DQA1* SNP genotypes were well clustered upon inspection of the respective cluster plots, and therefore it is unlikely that the genotypes were miss-called.

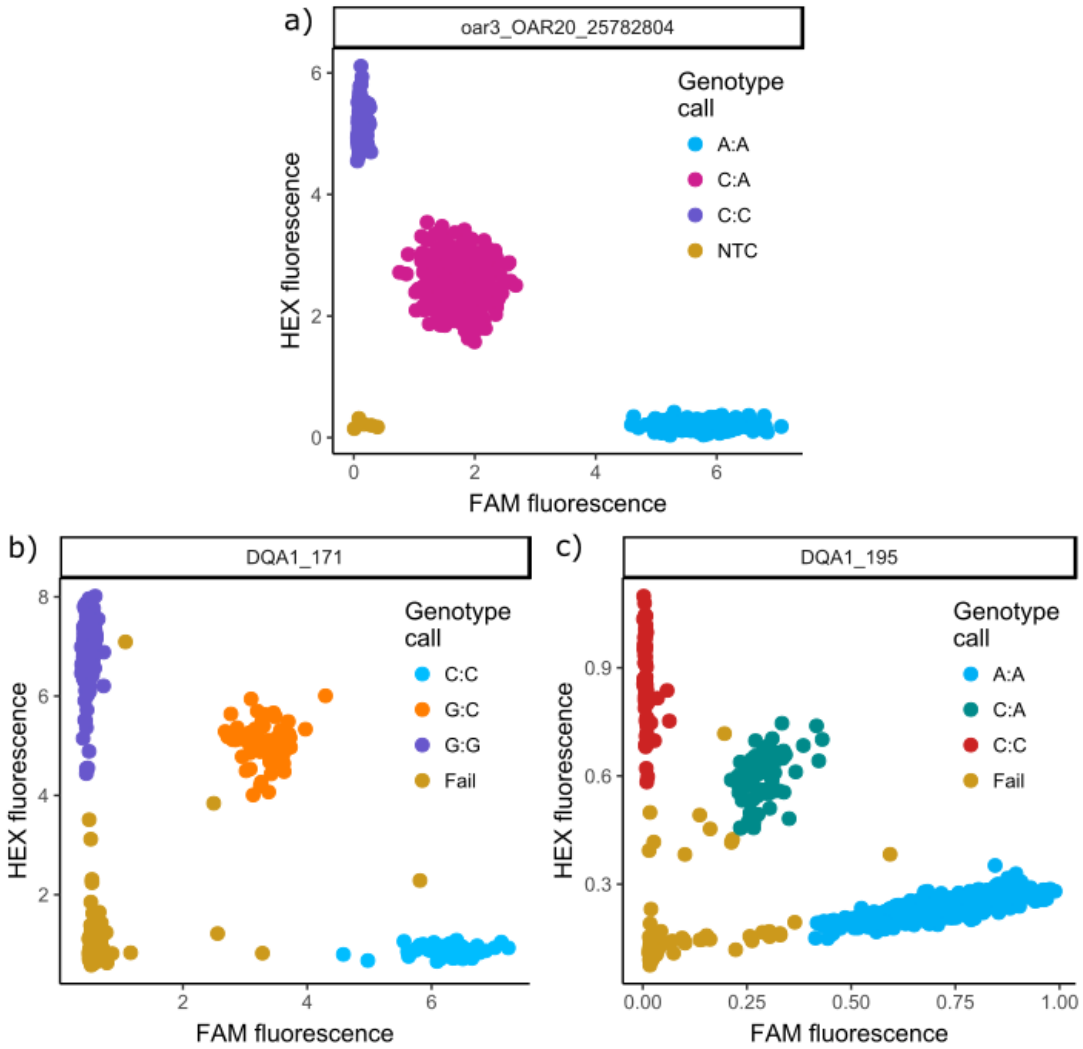



Figure 3.7. Cluster plots showing genotype calls for a) a typical example of a SNP genotyped by KASP, showing well-clustered genotypes that are easily called automatically by LGC software. NTC = No template control. *DQA1* SNPs, b) and c), were not as well clustered, and fail genotypes include both failed amplifications and the absence of amplification for *DQA1*null* homozygous individuals. Homozygous genotype calls (A:A, C:C and G:G) include both true homozygotes (two copies of the allele) and heterozygotes with one haplotype which is *DQA1**null.

Table 3.4. Conflicting expected and actual *DQA1* genotypes for five individuals.

| ID | SNP diplotype | Expected DQA1_171 | Actual DQA1_171 | Expected DQA1_195 | Actual DQA1_195 |
|---|---|---|---|---|---|
| 4528 | A/G | Fail | Fail | Fail | A:A |
| 4553 | A/A | Fail | Fail | Fail | A:A |
| 6157 | A/D | G:G | C:C | A:A | C:C |
| 6272 | A/E | Fail | G:G | Fail | A:A |
| 6543 | A/G | Fail | G:G | Fail | A:A |

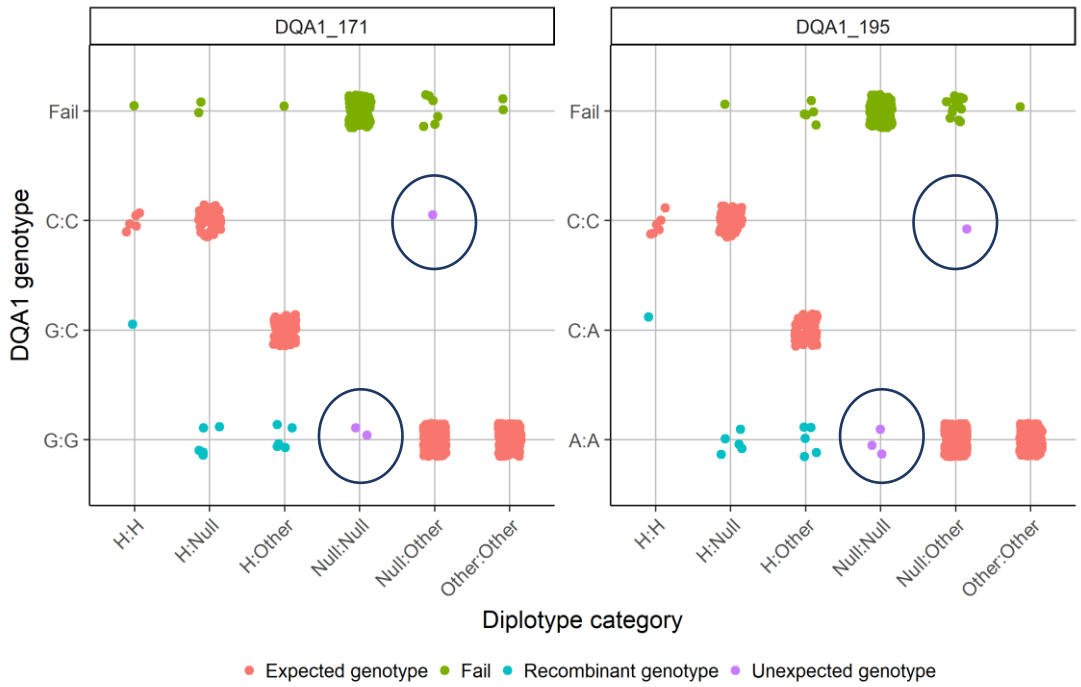



Figure 3.8. Comparison of the SNP diplotype category with *DQA1* genotypes to detect recombinant B to H SNP haplotypes. Haplotypes are grouped as H, which is involved in the recombination event, or according to whether or not they carry a DQA1 locus (Other includes haplotypes B, C, D and F, and Null includes A, E and G). Five individuals had unexpected DQA1 genotypes (circled), either Null:Null or Null:Other haplotype combinations. Points are shown jittered to indicate samples sizes.

## 3.5 Discussion

Analyses of the evolutionary processes underlying variation at the MHC require population-scale data. To achieve this, a SNP-based genotyping method was developed for rapid and accurate determination of the class IIa diplotypes in Soay sheep. Using SNPs selected from the Ovine Infinium HD BeadChip, a linkage region surrounding the class IIa genes was identified, and from it a minimal panel of 11 SNPs was selected that would enable imputation of the class IIa haplotypes. Furthermore, a recombination event between haplotypes B and H was detected, and two intragenic SNPs were selected to detect such recombinant haplotypes. Finally, this panel of 13 SNPs was tested in 801 animals using KASP genotyping technology.

### 3.5.1 Representation of MHC class IIa on 50K and HD SNP chips

To my knowledge, this is the first time that MHC class II haplotypes or alleles have been detected using a SNP chip on a non-human species. The class IIa region was very poorly represented on the 50K SNP chip, with only a single SNP located in the class IIa region and it was therefore unable to capture class IIa variation. However, 31 SNPs on the HD SNP chip were identified within the peripheries of and flanking the class IIa region, and were representative of the class IIa haplotypes.

### 3.5.2 Linkage disequilibrium within the class IIa

Pairwise measures of $D'$ and $r^2$ were unable to detect any variation in LD which would enable identification of SNPs associated with the class IIa genes. Pairwise $D'$ was typically very high between 24.5 and 26.5 Mb on chromosome 20, whilst pairwise $r^2$ was typically very low in this region. This opposing pattern of $D'$ and $r^2$ measures of LD was also the typical pattern identified across the class IIa, IIb and III regions in Rylington Merino sheep (Lee *et al.* 2012). VanLiere and Rosenberg (2008) show that $r^2$ and $D'$ are not always directly related, and $r^2$ is sensitive to MAF, especially MAF < 0.3 which is not uncommon within the MHC (see Table 3.3). Lee *et al.* (2012) showed that, in Rylington Merino sheep, pairwise LD using $D'$ was high within class IIa but decreased between the class IIa and class IIb regions, as well as between class IIa and class III regions. This suggests recombination *within* the class IIa is reduced compared to recombination between it and the neighbouring MHC subregions. Pairwise LD measures were therefore perhaps unlikely to reveal a relationship between the class IIa haplotypes and SNPs on the HD chip.

An alternative method of building up a picture of LD between the SNPs and the class IIa region was used where by using the four gamete rule was used to identify linkage blocks and use these were used as a starting point to cumulatively add SNPs to identify where degradation haplotypes degraded. This detected a linkage region of 31 SNPs that matched the class IIa haplotypes (Figure 3.4). This class IIa linkage region extended between 25.39 – 25.81 Mb according to the allocated physical positions of the SNPs, which sits squarely within the estimated class IIa gene region of 25.35 – 25.81 Mb. The downstream extent of the class IIa gene region is marked by the gene *BTNL2* (Table 3.1), which because it is a single copy gene is more likely to be accurately mapped than the other multi-copy genes in this region. The overlap between the class IIa gene region and the SNP linkage region suggests that at least 24 of the 31 SNPs are located within

the class IIa. Whether the SNPs are located within coding regions, however, cannot be determined due to the poor mapping of the DQ genes.

### 3.5.3 Recombination between B and H haplotypes

A recombination event between haplotypes B and H was identified in one individual (ID 4179) genotyped on the HD SNP chip. Sanger sequencing of the class IIa genes of 4179 revealed haplotype B, however the SNP haplotype within the class IIa linkage region was H (denoted H*). The recombination event must have occurred downstream of the expressed class IIa genes characterised within the Sanger haplotypes (*DRB1, DQA* and *DQB*) but upstream of SNP oar3_OAR20_25742763, and therefore the functional class IIa haplotype was unaffected. Individuals included on the HD SNP chip are highly representative of genetic variation within the Soay sheep population (Johnston *et al.* 2016), in the case of 4179 because this male had a large number of descendants. This recombinant haplotype event could, therefore, be expected to occur within the population at a low, but detectable frequency and especially in the descendants of 4179.

### 3.5.4 Development of a fast class IIa haplotype imputation method

There were 31 HD SNPs within the peripheries of and flanking the class IIa gene region that were capable of identifying all eight Sanger haplotypes, but not the recombination between haplotypes B and H. All 31 HD SNPs passed KASP *in-silico* validation, and, using only 11 of these HD SNPs, it was possible to impute the Sanger haplotypes whilst including degeneracy to allow for genotyping failures. The recombinant haplotype H* was rendered visible by using two intragenic *DQA1* SNPS.

### 3.5.5 KASP haplotyping using 11 HD SNPs

The final panel of 11 HD SNPs was genotyped at > 5 loci in 801 out of 941 individuals, with a genotyping success rate of greater than 94.5 % for all loci. There were 140 individuals which were excluded because more than 50 % of loci failed, which was likely due to poor DNA quantity or quality. The DNA plates used for KASP genotyping were the same as previously used for the 50K SNP chip, and so DNA may have been exhausted or possibly somewhat degraded by freeze-thawing.

HD SNP haplotypes from 799 of the 801 individuals matched the eight haplotypes previously called, plus two novel haplotypes. Comparison of 37 individuals genotyped

by both KASP and HD SNP chip methods revealed one haplotype incorrectly phased by the KASP SNPs (individual 4399). The programme BEAGLE (Browning & Browning 2007), which was used to phase the SNPs into haplotypes, uses a hidden Markov model to determines the most likely haplotypes for all individuals, which incorporates information on the haplotypes known to exist within the population to impute the haplotypes of individuals with missing genotypes. Therefore, as the number of missing genotypes increases, so does the probability of incorrect imputation. Low genotyping rate (7 out of 11 loci) likely caused the incorrect phasing of one haplotype for individual 4399. The two novel haplotypes may be artefacts caused by genotyping errors or incorrect phasing, or may be genuine variants such as rare haplotypes, previously undetected recombinant haplotypes, or due to a mutation at a SNP site. It will therefore be important to apply a higher genotyping rate in subsequent analyses and consider using additional quality control measures, such as checking for Mendelian inheritance of haplotypes.

Indirect MHC typing methods, such as SNP genotyping, are faster and cheaper than direct sequencing methods, but may incur an increased error rate which must be minimised. Genotyping errors can occur at different stages including SNP genotyping, SNP phasing and MHC haplotype imputation. At the genotyping stage, genotyping errors and allelic dropout can produce inaccurate genotypes. Phasing SNPs to determine the two SNP haplotypes carried by an individual, carried out here using BEAGLE, uses information from all individuals within the dataset to calculate the most likely haplotypes. BEAGLE also uses this information to impute missing SNP genotypes. Therefore, phasing errors are increased when there is a high genotyping error rate. Imputation of MHC haplotypes was carried out here by comparing SNP profiles for known class IIa haplotypes with those phased by BEAGLE. It is therefore possible errors to build up with each stage from genotyping, to phasing to MHC haplotype imputation. Thus, by using a more stringent genotyping rate cut-off across all SNPs, genotyping errors can be limited. Additionally, the extensive Soay sheep pedigree enables Mendelian inheritance checks to be carried out, which limits potential genotyping or phasing errors from being carried through and utilised in the final dataset.

### 3.5.6 *DQA1* intragenic SNPs

Two SNPs within exon 2 of *DQA1* were used to identify the recombinant haplotype by detecting a mismatch between the haplotype imputed by the 11 HD SNPs and the class IIa genes. A null allele (*DQA1*null*) occurs at the DQA1 locus, and the presence of null alleles for any genotyping assay is less than ideal as it is not possible to separate a null genotype from a failed assay. However, DQA1_171 and DQA1_195 were the only intragenic SNPs to pass the *in-silico* KASP validation. Despite the null alleles, in all cases of an individual carrying haplotype H, as imputed by the 11 HD SNPs, DQA1 genotypes were as expected for haplotype H or as predicted for a recombinant haplotype B.

When selecting intragenic SNPs, the potential for the *DQA1* primers to bind to *DQA2* or *DQA2-like* loci was not considered during *in-silico* validation. This is because all potential intragenic SNPs would have been rejected due to the level of polymorphism in flanking regions. When alleles for all *DQA* loci are aligned (Appendix D), it is clear that there may be some potential for KASP primers to bind to *DQA2* or *DQA2-like* loci (note that the actual primer sequences used by LGC are unknown). Alternatively, there may be pseudogenes which the primers could bind to and amplify. It seems unlikely that mis-priming to *DQA2/2-like* loci could explain the amplification of DQA1_171 in *DQA1*null* homozygous individuals as this would result in a C:C genotype, yet G:G genotypes were observed, though it could explain the amplification at DQA1_195. Erroneous SNP haplotype imputation could also explain the unexpected *DQA1* genotypes and may be the reason for the conflicting *DQA1* genotypes of individual 6157 whose imputed diplotype was based on only 7 of the 11 HD SNPs. Due to this potential for cross-amplification to either non-target DQA loci or pseudogenes, it would be advisable to use the DQA1 loci to detect only the known recombination between B and H haplotypes.

Whilst the *DQA1* SNPs were effective for detecting the identified downstream B to H recombination, it is unlikely to be possible to detect novel recombination events involving upstream SNPs between haplotypes A and G. The six SNPs located upstream of the class IIa genes and the *DQA1* SNPs were identical for these this haplotype pair (see Figure 3.4 and Figure 3.5). All other haplotype pairs have unique upstream and downstream SNP profiles. The B to H recombination was identified in individual 4179 because he was highly representative of genetic variation in the Soay pedigree and was therefore typed at the HD SNP chip (Johnston *et al.* 2016) due to having a large number of descendants and hence the value of being able to identify this recombination event.

Recombinations between haplotypes A and G could occur and not be detected using this SNP panel, but it would be unlikely to affect large numbers of individuals or it probably would have been detected in the individuals typed for the HD SNP chip.

### 3.5.7 Conclusion

A rapid method for inferring the diplotypes of the MHC class IIa region in the Soay sheep population was developed using KASP genotyping. Using genotypes from 188 individuals, a panel of 11 SNPs were selected from the Ovine Infinium HD BeadChip to infer the class IIa haplotypes. Additionally, two intragenic *DQA1* SNPs were selected to detect a recombination between haplotypes B and H. This panel of 13 SNPs was validated using KASP in 801 individuals, proving capable of inferring the class II diplotypes for all but three individuals. Additional quality control measures should be refined to ensure high quality class IIa haplotype imputation.

# 4. SNP genotyping to impute MHC class IIa haplotypes in a large sample of Soay sheep and evidence of balancing selection

## 4.1 Abstract

In chapter 3, a panel of 13 KASP SNPs was selected to impute the MHC class IIa haplotype. In this chapter, the panel of SNPs was genotyped in a further 5951 Soay sheep. Stringent QC measures were applied, including testing for systematic errors, comparison of genotypes of a repeatedly typed individual, exclusion of samples with more than four genotyping errors, and Mendelian inheritance checks between the offspring and parental genotypes or haplotypes. After applying these QC measures, class IIa haplotypes were inferred for 5349 individuals, with a SNP genotyping error rate of 0.48 % in the repeatedly typed individual. Twenty novel haplotypes identified in 27 individuals were excluded from further analyses as they probably arose due to SNP genotyping errors. The recombinant haplotype (from B to H) was identified in 56 individuals using the two intra-genic *DQA1* SNPs, and its t origin and inheritance traced through the pedigree. Finally, haplotype and diplotype frequencies were assessed at different life history stages and within the standing population each year between 1985 and 2012. The Ewens-Watterson test showed evidence of balancing selection at all life history stages and standing populations considered. No evidence of deviation from Hardy-Weinberg equilibrium was observed at any life history stage or in any standing population.

## 4.2 Introduction

Investigating how evolution acts to maintain the high levels of polymorphism requires well-characterised MHC data for large numbers of individuals (Apanius *et al.* 1997; Spurgin & Richardson 2010). Accurate, complete and locus-specific MHC genotypes are necessary to explicitly investigate the different modes of balancing selection that may be acting, in particular, to distinguish heterozygote advantage from negative frequency-dependent selection or frequency dependent selection (Spurgin & Richardson 2010; see Chapters 1 and 2). Statistical analysis of highly polymorphic loci, such as the MHC, requires large sample sizes to generate statistical power and is enhanced in the presence of rare alleles (Apanius *et al.* 1997; Hill 1998). Rare alleles are of particular interest in studies of the MHC as they may confer increased fitness under negative frequency-dependent selection or in heterozygote form under heterozygote advantage. Due the nature of their rarity, detection of rare alleles requires large sample sizes.

Sequence-based genotyping of MHC alleles is accurate and reliable for most loci in sheep (Ballingall & Tassi 2010; Ballingall *et al.* 2017). However, Sanger sequencing of individual loci is time-consuming and expensive, prohibiting the genotyping (or haplotyping) of large numbers of individuals. In Chapter 3, an alternative method of determining class IIa haplotypes in Soay sheep was developed which enables high-throughput haplotyping of large numbers of individuals. The panel included 11 SNPs selected from the Ovine Infinium HD chip which were located in the peripheries of the class IIa region and could impute the eight class IIa haplotypes identified in chapter 2, with the addition of two *DQA1* intragenic SNPs to identify a recombinant SNP haplotype.

This panel of SNPs was tested in 941 sheep in chapter 3, and the imputed SNP haplotypes were compared to Sanger haplotypes in 37 individuals. The haplotype was inferred incorrectly for a single individual due to genotyping failures at four of the 11 SNP loci. This indicated that improved quality control (QC) measures were needed to obtain impute MHC class IIa haplotypes accurately. In this chapter, the KASP SNP panel was used to genotype an additional 5010 sheep and describe the quality control measures implemented.

After generating a data set of high quality MHC class IIa haplotypes for a large sample of Soay sheep, some basic analyses of the diplotype frequencies were conducted. The genotyped sheep were mostly born between 1985 and 2012, enabling diplotype frequency changes to be considered over time. Second, if heterozygote advantage (Doherty & Zinkernagel 1975) or negative frequency-dependent selection (NFDS; (Penn 2002) were operating to maintain diversity within the population, an excess of heterozygotes would be predicted under certain circumstances. Under heterozygote advantage, selection would be acting against homozygotes, whereas under NFDS, rare alleles are more likely to occur in heterozygous form (Apanius *et al.* 1997; Spurgin & Richardson 2010) and therefore selection will favour heterozygotes carrying the rare allele.

Evidence of recent balancing selection can be tested using the Ewens-Watterson homozygosity test for neutrality (Ewens 1972; Watterson 1978). This test compares observed homozygosity to the expected homozygosity under neutrality for the given number of alleles in a given sample size. Expected homozygosity is calculated according to Hardy-Weinberg proportions. Observed levels of homozygosity that are significantly lower than expected demonstrates that allele frequencies are more even than would be

expected under neutrality, invoking balancing selection as an explanation. Levels of homozygosity that higher than expected under neutrality, on the other hand, indicates directional selection (Garrigan & Hedrick 2003).

Hardy-Weinberg equilibrium (HWE) predicts genotype (or diplotype) frequencies from the allele (or haplotype) frequencies for a population that is mating randomly at loci with no selection, mutation or migration (Guo & Thompson 1992). Therefore, evidence that the actual diplotype frequencies deviate from those predicted under HWE indicates that natural selection (including heterozygote advantage and NFDS) or sexual selection (including MHC based mate choice) may be acting on the haplotypes (Hedrick 2004), or that some other demographic process acting (Guo & Thompson 1992).

Different modes of selection, for example parasite-mediated selection (PMS) and pre- or post-copulatory sexual selection could affect haplotype and diplotype frequencies at different life history (LH) stages. MHC-based mate choice (either pre- or post-copulatory) would be expected to affect the frequencies in conceived individuals. PMS, on the other hand, would only be expected to act after birth, and if gastro-intestinal nematodes (GIN) were a major driving force of PMS, only after lambs become exposed to the larvae on the pasture. In the Soay sheep, selection due to GIN (as measured by August faecal egg count) is only significant in lambs and manifests as higher first winter mortality of heavily infected lambs (Hayward *et al.* 2014a), and there is appreciable yearling mortality (Clutton-Brock *et al.* 2004a), so haplotype frequencies might be expected to change following first and second winter mortality. In this chapter, the Ewens-Watterson test and deviations from HWE were assessed at different early life history (LH) stages, and within the standing population (defined here as the individuals known to be alive in August of a given year).

## 4.3 Methods

### 4.3.1 KASP genotyping

The DNA used for KASP genotyping had been previously extracted, plated in 96 well plates and normalised (originally to 50ng/ul) for SNP genotyping on the Illumina Infinium ovine 50K SNP array (Bérénos *et al.* 2014). The DNA was extracted by Camillo Bérénos and Phil Ellis using the Qiagen DNEasy kit in either tube or plate format

following manufacturer's protocol, except that final elution was in 2 x 50ul of elution buffer in order to obtain high DNA concentration. Most DNA was extracted from ear punch samples collected from neonatal lambs or at later first capture, but for some individuals, DNA was extracted from blood samples from live captures or post mortem tissue (ear or muscle). Some samples were from foetuses from pregnant ewes that died in winter. There were 66 DNA plates, including the 10 genotyped and analysed in chapter 3. One individual, the 'golden sheep' (ID 7568), was included in a random well on plate 3 and plates 6 – 66 to confirm genotyping consistency. Kompetitive Allele-Specific PCR (KASP) genotyping was carried out by LGC Genomics (Hoddesdon, UK) for 5951 individuals at 13 MHC loci on chromosome 20, as detailed in chapter 3 as well as 11 non-MHC SNPs selected by other project members for non-MHC studies.

### 4.3.2 Quality control

A flow-chart detailing the bioinformatic quality control (QC) workflow is shown in Figure 4.1. QC was carried out using 22 non-intragenic SNPs (both MHC and non-MHC), but not the two class II intragenic loci DQA1_171 and DQA1_195. The intragenic *DQA1* SNPs were excluded from QC analyses as these loci are effectively tri-allelic – they have a major allele, a minor allele and a null allele (DQA1 locus absent from haplotype); consequently, Mendelian inheritance of ascertained genotypes is not expected and a null genotype is valid. In the first round of quality control, any individuals with 11/22 or more missing genotypes (≥ 50%) were excluded as this suggested poor DNA quality or quantity, which would affect subsequent QC analyses.

In the next stage, repeat genotypes for individuals were compared. The genotypes of the golden sheep were compared to identify any systematic patterns of genotyping errors, such as by plate or by SNP locus.

A number of other individuals were genotyped more than once for reasons unrelated to this work (e.g. when DNA was used for genotyping on the Illumina 50K Beadchip, the DNA failed at the first attempt and so was repeated on a subsequent plate – note that differences in genotyping technologies between Illumina and KASP may mean that a sample could be successful for one method but not the other). These repeated samples were not compared as a measure of genotyping error as the DNA was expected to be low quality. Genotyping attempts for a repeated individual were compared with parental genotypes. The retained set of genotypes was that with the fewest missing loci, unless this sample had a greater number of Mendelian mismatches to the sire or dam,

in which case the sample with the lower number of mismatches was kept. A preliminary Mendelian inheritance check was then performed for all samples using the Soay sheep pedigree (Bérénos *et al.* 2014).

Following this initial round of quality control, samples on plates 1-3 were excluded due to high rates of allelic dropout and genotyping errors, including in the golden sheep (see results). In the second round of QC, above QC stages were repeated from the start excluding plates 1 to 3. This was to prevent genotypes from plates 1 to 3 from any repeated individuals impacting the paternal mismatch checks.

Next, PLINK v1.90 (Chang *et al.* 2015) was used to convert the data to vcf format (as required for BEAGLE), and to exclude individuals with four or more missing genotypes (missing rate greater than 0.14). As described in chapter 3, samples analysed with four or more missing genotypes can lead to inaccurate phasing. At this stage, the non-MHC SNPs were discarded by excluding all SNPs on chromosomes other than 20.

#### 4.3.2.1 MHC haplotype phasing and Mendelian inheritance check

Phasing was carried out using BEAGLE 4.0 (Browning & Browning 2007) with standard settings, and parsed using a custom script in R (S. E. Johnston, unpub.). Haplotypes were matched to haplotypes identified in chapter 3 and assigned alphabetical nomenclature (A to H), and novel haplotypes were assigned NA.

A Mendelian inheritance check was carried out on the inferred haplotypes for each individual using the Soay sheep pedigree (Bérénos *et al.* 2014). Individuals whose phased MHC diplotypes mismatched the diplotype of at least one parent were excluded from further analysis.

### 4.3.3 DQA1 loci

The methodology established in chapter 3 was used to identify SNP haplotype B to H recombinants. Because *DQA1* genotypes for individuals homozygous for *DQA1*null* haplotypes were found to be somewhat unreliable in chapter 3, *DQA1* genotypes were only analysed for individuals carrying at least one H haplotype. Genotypes were assigned as H:H (both haplotypes are H), H:null (the alternative haplotype does not carry a *DQA1* locus so is *DQA1*null*) or H:other (the alternative haplotype does carry a *DQA1* locus), and the *DQA1* observed genotypes were compared to the expected genotypes. Individuals which were identified as carrying the non-intragenic SNP profile of haplotype H but the intragenic *DQA1* genotypes of haplotype B were temporarily

recoded as recombinant. A Mendelian inheritance check was then carried out to confirm that these recombinant haplotypes were inherited, rather than genotyping errors. Recombinant individuals were subsequently recoded as haplotype B.

### 4.3.4 Ewens-Watterson test and HWE

The Ewens-Watterson test and deviations from HWE were investigated in different LH stages and in the standing populations of each year. The LH stages considered here were: known conceived, live born, survived to first August (approx. 4 months) and survived to second August (approx. 16 months). Individuals were included in the 'known conceived' category if they were live born, still born, or foetuses obtained from mothers who died over winter. Individuals were included in the 'live born' category if they survived at least until tagging which typically occurs within the first week of life. 'Live born' also includes individuals that were not tagged as lambs, for example because they were born after the lambing team left the island, but were caught and sampled subsequently, for example during the summer catch. 'Survived to first August' and 'survived to second August' includes individuals known to be alive in their first and second Augusts, respectively.

The standing population in any given year was all individuals known to be alive in August of that year. Individuals were known to be alive if they were observed in the summer census or August catch, or if they had a known birth year prior to the given year and were observed alive or known to have died in a later year.

The Ewens-Watterson homozygosity test of neutrality (Ewens 1972; Watterson 1978) was calculated for each LH stage (though not by cohort) and for the standing population in each year, and was performed in PyPop v0.7.0 (Lancaster *et al.* 2007). This test compares the observed homozygosity statistic $F_{obs}$ (sum of the squared allele – here haplotype – frequencies calculated under Hardy-Weinberg proportions) to the expected homozygosity statistic $F_{exp}$ (the homozygosity expected under neutrality). $F_{nd}$ is the normalized deviate of the homozygosity ($F_{obs}$ minus $F_{exp}$, divided by the square root of the variance of the expected homozygosity), and significant negative values of $F_{nd}$ indicate balancing selection as the level of observed homozygosity is lower than expected. If $F_{nd}$ were significantly positive, this would imply directional selection. Slatkin's exact *P* values (Slatkin 1994, 1996) were calculated using a Markov-chain Monte-Carlo implementation, which is the probability of observing the $F_{exp}$ under neutrality that is less than or equal to $F_{obs}$ for the given sample size and number of

alleles. Sequential Bonferroni correction (Holm 1979) was used to account for multiple testing.

HWE tests were performed in the R package GENEPOP v1.0.5 (Rousset 2008) as probability tests (the exact test of Haldane 1954; Guo & Thompson 1992; Weir 1996) where the null hypothesis is that the same allele frequencies are observed) and *U* tests for heterozygote excess (where the alternative hypothesis is heterozygote excess Raymond & Rousset 1995) for all LH stages considering all year cohorts as a single population, and to identify if there were any deviations from HWE within particular years, the same tests were carried out for each cohort (LH stages) or year (standing population). The single population test was not carried out for the standing population as many individuals were present in multiple years. Sequential Bonferroni correction (Holm 1979) was used to account for multiple testing.

Unless otherwise stated, all analyses were carried out in R version 3.3.3.

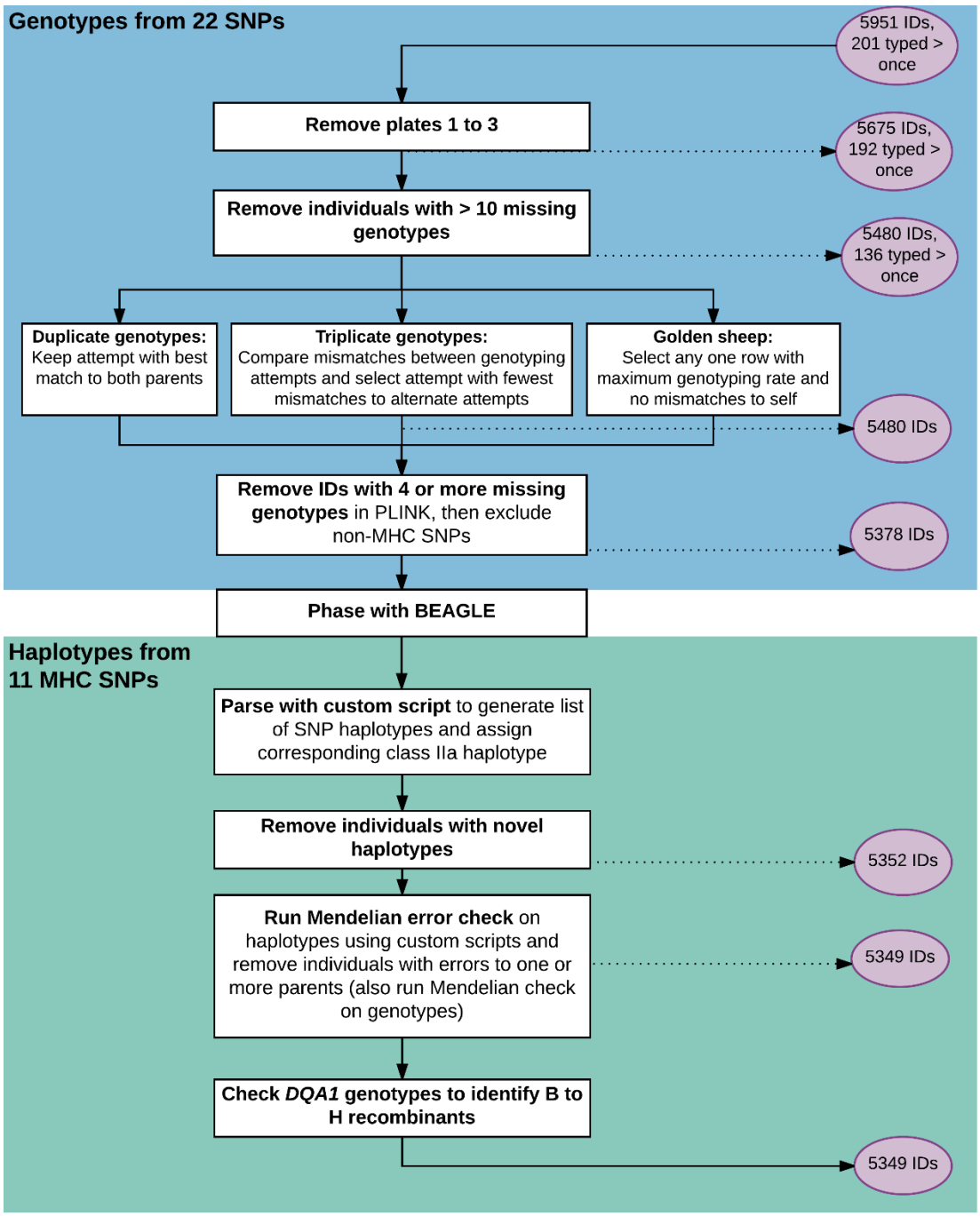



Figure 4.1. Flow chart illustrating the bioinformatic workflow for QC of KASP genotypes, showing processes in rectangles and the number of individuals at each stage in circles. QC processes using the genotypes from 22 SNPs are shown in the top blue box, and using the haplotypes phased from 11 MHC SNPs in the bottom green box. Note that this represents the second round of the bioinformatic workflow after the first identified poor quality genotypes across plates 1-3.

# 4.4 Results

## 4.4.1 Quality control

### 4.4.1.1 Round one – identification of poor quality on plates 1-3

There were 5951 individuals genotyped by KASP (see Figure 4.1 for sample sizes throughout QC). The first round of Mendelian inheritance checking found 285 individuals that had at least one genotype mismatch to one or both parents. Of these, 193 individuals were located on DNA plates 1-3 (Figure 4.2). One or both parents of an additional 31 individuals with mismatches were located on plates one to three. Therefore, 224 of the 285 individuals with mismatches were genotyped from three DNA plates. This strongly suggested that there were DNA quantity or quality problems with plates 1-3. Furthermore, the genotypes of all individuals with genotype mismatches to their parents were homozygous, suggesting that allelic dropout was occurring.

The golden sheep 7568 was genotyped on 62 of the 66 DNA plates, including plate 3 (but not 1 or 2). Genotypes were obtained for all 22 loci for the golden sheep on plate 3, but nine genotypes were inconsistent with the profile for that individual obtained from all other plates. These incorrect genotypes include both allelic dropout (i.e. a heterozygous locus appeared homozygous) and additional alleles (i.e. individual appeared heterozygous but true genotype was homozygous).

Both the preliminary Mendelian error checks and comparison of the golden sheep profiles suggest that plates 1-3 had high genotyping error rates and so they were deemed unreliable. Plates 1-3 were excluded from further analyses.

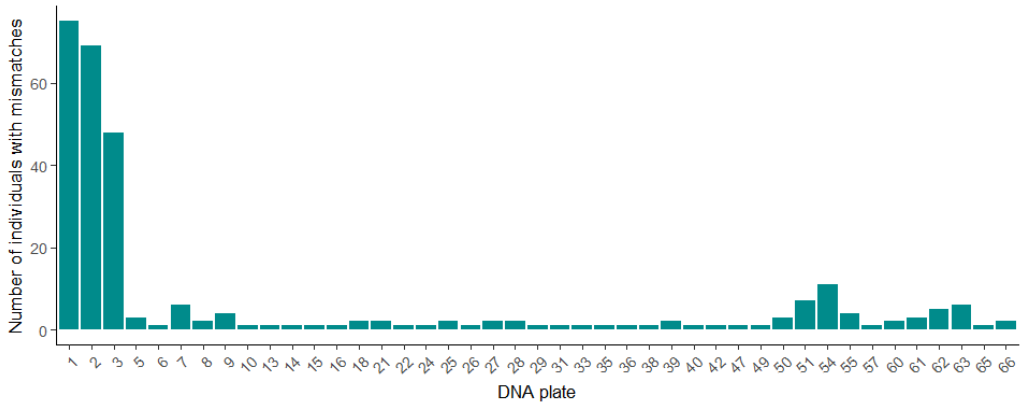



**Figure 4.2.** The DNA plates on which individuals with at least one genotype mismatch to parents were located. DNA plates which did not have any individuals with mismatches are not shown.

#### 4.4.1.2 Round two – after excluding plates 1-3

After excluding plates 1-3, all 22 genotypes for 5675 individuals were run through the QC workflow again. There were 195 individuals excluded due to having more than 50 % missing genotypes. Of the remaining 5480 individuals, 136 were genotyped more than once. Genotypes were obtained for 57 out of 62 genotyping attempts for the golden sheep and errors were found on four plates (Table 4.1); of these, one profile would have been excluded during the second round of quality control due to high missing genotype rates (nine genotypes missing on Plate 10). The remaining three profiles had errors at one (plate 65) and two loci (plates 7 and 8). This represents an overall genotyping error rate of 0.48%, or a maximum of 1.85 % per locus.

After retaining one genotype for 5480 individuals and removing repeat genotypes for other repeated individuals, the 22 SNP genotypes were run through PLINK, 102 individuals were removed due to having four or more missing genotypes, and then the non-MHC SNPs were excluded. The remaining 5378 individuals were then phased by BEAGLE using the 11 MHC SNPs, revealing eight common haplotypes, which exactly matched the eight class IIa haplotypes identified in chapter 3, and an additional 20 novel haplotypes found in 26 individuals (Table 4.2).

Mendelian inheritance checks were then carried out on both the genotypes for each of the 11 MHC loci and the phased haplotypes (including 3971 maternal and 3711 paternal comparisons). This identified 14 mismatches at the genotype level (Table 4.3), of which two offspring and nine parents were found to carry novel haplotypes. At the haplotype level, mismatches were found between 2 offspring carrying a novel haplotype and their parents, and 11 between offspring and parents who carried novel haplotypes. An additional three individuals were identified with mismatches to parents involving non-novel haplotypes (Figure 4.3). Individual 4597 (Figure 4.3a) mismatched both parents. Individual 5864 (Figure 4.3b) was found to have a haplotype mismatch to her only known parent. Individual 8450 (Figure 4.3c) mismatched his mother's haplotype. The father of 8450 was excluded prior to the Mendelian inheritance check due to carrying a novel haplotype. Nevertheless, 8450 also mismatched his mother's genotype. In all three cases, there is no good evidence to suggest that the known parental genotypes were wrong, suggesting the error is more likely to have occurred in the offspring genotype. Therefore, these three individuals (4597, 5864 and 8450) were excluded from further analyses.

Table 4.1. Repeat genotypes of the golden sheep (ID 7658) for each SNP, of a total of 57 genotyping attempts (not including plate 3). The 11 MHC SNPs are shown first (ID beginning oar3), followed by the non-MHC SNPs. Note that 5 genotyping attempts for the golden sheep were complete failures.

| SNP ID | Correct genotype | Number of correct genotypes | Number of genotyping errors (incorrect genotype) | Plate of genotyping error | Error rate per SNP (%) |
|---|---|---|---|---|---|
| oar3_OAR20_25393230 | A:A | 56 | | | 0 |
| oar3_OAR20_25417300 | A:A | 55 | 1 (G:G) | Plate 10 | 1.78 |
| oar3_OAR20_25419505 | C:C | 54 | | | 0 |
| oar3_OAR20_25441205 | T:C | 53 | 1 (C:C) | Plate 7 | 1.85 |
| oar3_OAR20_25447962 | G:A | 54 | | | 0 |
| oar3_OAR20_25458354 | G:G | 56 | | | 0 |
| oar3_OAR20_25742711 | A:A | 55 | 1 (G:A) | Plate 7 | 1.78 |
| oar3_OAR20_25742763 | T:T | 57 | | | 0 |
| oar3_OAR20_25759002 | G:A | 53 | 1 (G:G) | Plate 8 | 1.85 |
| oar3_OAR20_25761264 | C:A | 56 | 1 (C:C) | Plate 8 | 1.75 |
| oar3_OAR20_25782804 | G:G | 54 | | | 0 |
| Agouti_indel | AGGAA:AGGAA | 57 | | | 0 |
| Agouti_SNP_Correct | T:T | 56 | | | 0 |
| rs398157763 | T:T | 57 | | | 0 |
| rs398989339 | G:G | 57 | | | 0 |
| rs408103852 | G:A | 55 | 1 (G:G) | Plate 65 | 1.78 |
| rs411614551 | T:C | 52 | | | 0 |
| rs413407592 | C:C | 56 | | | 0 |
| rs417406686 | C:C | 57 | | | 0 |
| rs428766742 | T:T | 56 | | | 0 |
| rs428943636 | G:G | 57 | | | 0 |
| Tyrp1 | T:G | 56 | | | 0 |

Table 4.2. Phased SNP haplotypes, their associated class IIa haplotype and the frequency of each haplotype. There were 20 novel SNP haplotypes at frequencies which did not match known class IIa haplotypes (denoted as NA). Lower case in the SNP profile for novel haplotypes indicates where the profile is one base different from a known class IIa haplotype.

| | SNP profile | Class IIa haplotype | Frequency |
|---|---|---|---|
| Known profiles | AACCGGATGCG | A | 1904 |
| | AACTAGGCAAA | B | 2337 |
| | AACTACATACA | C | 1233 |
| | AATTGGGCAAA | D | 317 |
| | AGCCGGGCAAA | E | 1041 |
| | GGCCGGGCACA | F | 1562 |
| | AACCGGATAAA | G | 1385 |
| | AACTAGATAAG | H | 949 |
| Novel profiles | AGCCGGGCAcA | NA | 4 |
| | AACCGGATAcA | NA | 3 |
| | AaCCGGGCACA | NA | 2 |
| | AACtGGATGCG | NA | 2 |
| | GGCCGGGCgCA | NA | 2 |
| | AACCGGgTGCG | NA | 1 |
| | AACTAGATAAa | NA | 1 |
| | AACTAGATAcA | NA | 1 |
| | AACTAGATgAG | NA | 1 |
| | AACTAGGCAcA | NA | 1 |
| | AACTAGGCgAA | NA | 1 |
| | AACtAGgTAAG | NA | 1 |
| | AACtGGATAAA | NA | 1 |
| | AAtCGGATAAA | NA | 1 |
| | AAtCGGATGCG | NA | 1 |
| | AGCCGGGCgAA | NA | 1 |
| | AgCTAGGCAAA | NA | 1 |
| | AgTCGGGCAAA | NA | 1 |
| | gACTAGGCAAA | NA | 1 |
| | GGtCGGGCACA | NA | 1 |

Table 4.3. Mendelian inheritance locus mismatches between offspring and parent. Offspring denoted by * were also subsequently found to have haplotype mismatches with their parents

| SNP ID | Offspring ID | Offspring genotype | Offspring haplotype | Parent | Parent ID | Parent genotype | Parental haplotype |
|---|---|---|---|---|---|---|---|
| oar3_OAR20_25759002 | 6716 | G:G | NA:A | Mother | 5466 | A:A | F:G |
| | 6916 | G:G | NA:A | Mother | 5662 | A:A | D:E |
| | 8433 | A:A | F:H | Father | 6789 | G:G | NA:A |
| | 8710 | A:A | F:H | Father | 6789 | G:G | NA:A |
| | 8803 | A:A | B:E | Mother | 6916 | G:G | NA:A |
| oar3_OAR20_25742711 | 4433 | A:A | A:H | Mother | 3564 | G:G | NA:E |
| | 4597* | A:A | A:H | Father | 4554 | G:G | E:F |
| | 5109 | A:A | C:H | Mother | 3564 | G:G | NA:E |
| | 6719 | A:A | G:H | Mother | 3564 | G:G | NA:E |
| | 8166 | A:A | A:H | Mother | 3564 | G:G | NA:E |
| oar3_OAR20_25417300 | 4597* | A:A | A:H | Father | 4554 | G:G | E:F |
| | 5864* | G:G | E:E | Mother | 3255 | A:A | A:B |
| oar3_OAR20_25782804 | 4597* | G:G | A:H | Mother | 4243 | A:A | E:G |
| | 4597* | G:G | A:H | Father | 4554 | A:A | E:F |

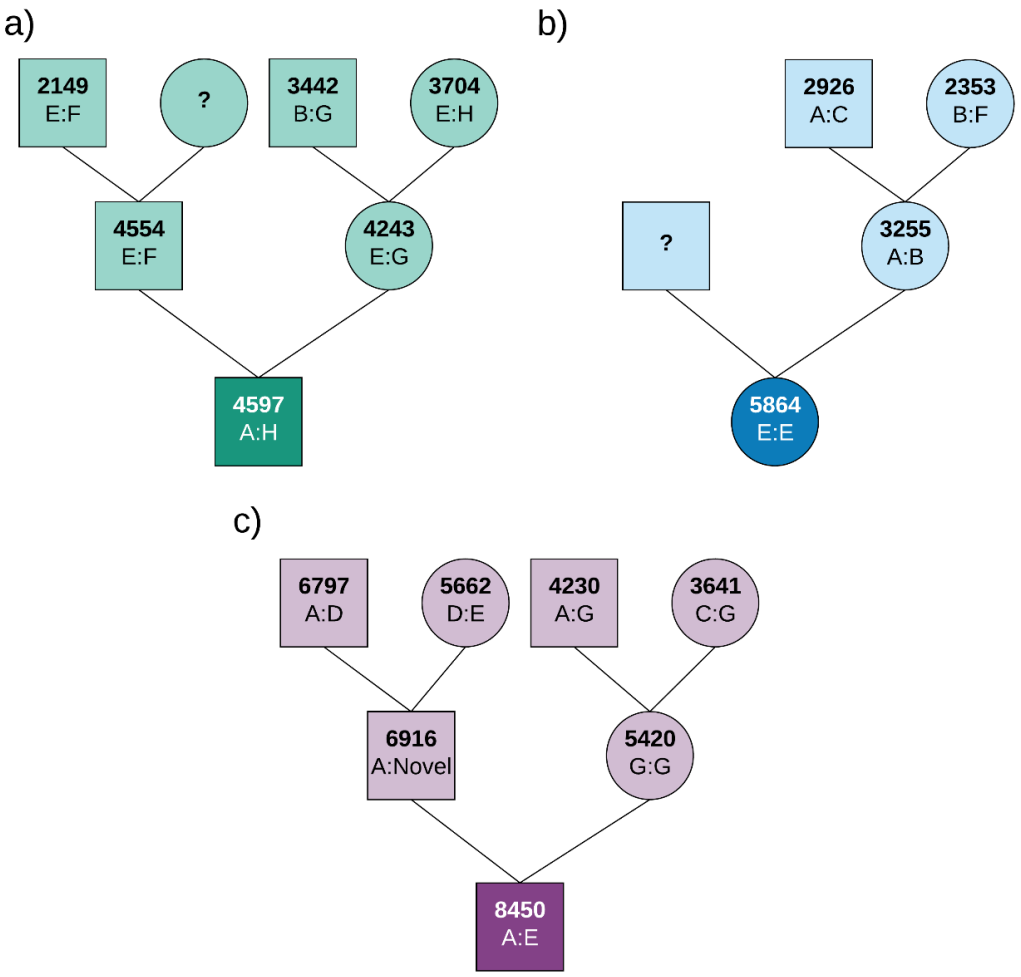



Figure 4.3. Pedigrees of three focal individuals that were identified as having haplotype mismatches to at least one parent. ID and inferred MHC haplotype are shown within the shapes (square is male, circle is female), and unknown parents are indicated by '?'. The focal individual is shown in the dark shapes at the base of the pedigrees.

### 4.4.2 DQA1 analysis to identify B/H recombinants

There were 56 individuals identified as carrying a B/H recombinant *DQA1* profile. Following recoding of these individuals as R (recombinant haplotype), only one individual (ID 4179) was identified as mismatching a parental genotype during the Mendelian inheritance check, mismatching the maternal genotype. This was the same individual that the recombinant haplotype was identified in when comparing Sanger to HD haplotypes in chapter 3. Of all the individuals carrying the B/H recombination, individual 4179 was born earliest and all subsequent individuals were found to be descendants of 4179 (Figure 4.4). Furthermore, haplotypes of the parents of 4179

(Figure 4.4) showed that his mother (3471) carried both B and H haplotypes, and thus the cross-over event that created the recombinant haplotype occurred during gametogenesis in 4179's mother. 4179's alternative haplotype (G) was consistent with direct Mendelian inheritance from his father (3341, A:G). Recombinant haplotypes were recoded as B.

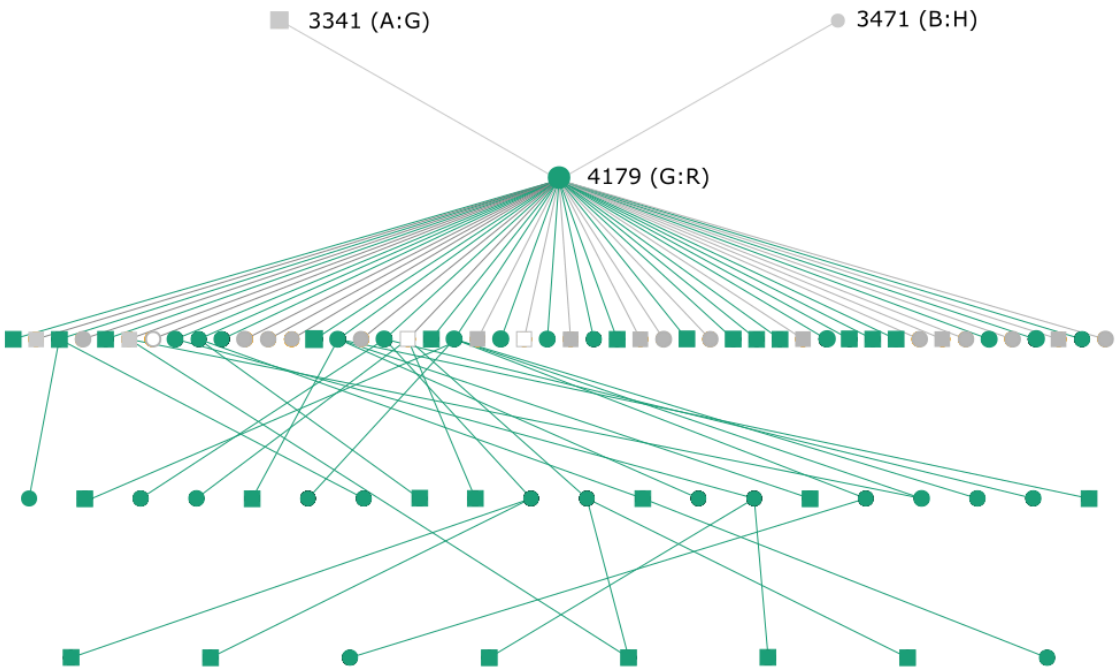



Figure 4.4. Pedigree of individual 4179. Green represents individuals carrying the recombinant haplotype, grey not carrying the recombinant haplotype, and open are individuals that have not been genotyped. Squares are males, circles are females. All offspring of ID 4179 are shown, but subsequent descendants are only shown if they carry the recombinant haplotype. Diplotypes of 4179 and his parents are shown in brackets.

### 4.4.3 Haplotype and diplotype frequencies

After excluding individuals rejected by quality control, removing individuals with novel haplotypes and recoding recombinant individuals, there were diplotypes for 5349 individuals, of which 17.9% were homozygous. Haplotype frequencies are shown in Figure 4.5a, and diplotype frequencies in Figure 4.5b. All possible diplotypes were observed.

Haplotype frequencies in individuals born in each 5-year period (of a total of 5137 individuals with known birth years) carrying each haplotype is shown in Figure 4.6. There was no visual evidence that within the timeframe considered (born between

1985 and 2012) there were major shifts in haplotype frequency, for example, no haplotype switches from rare to common. Figure 4.6 does suggest that over time there may be a decline in haplotype C and an increase in haplotype D.

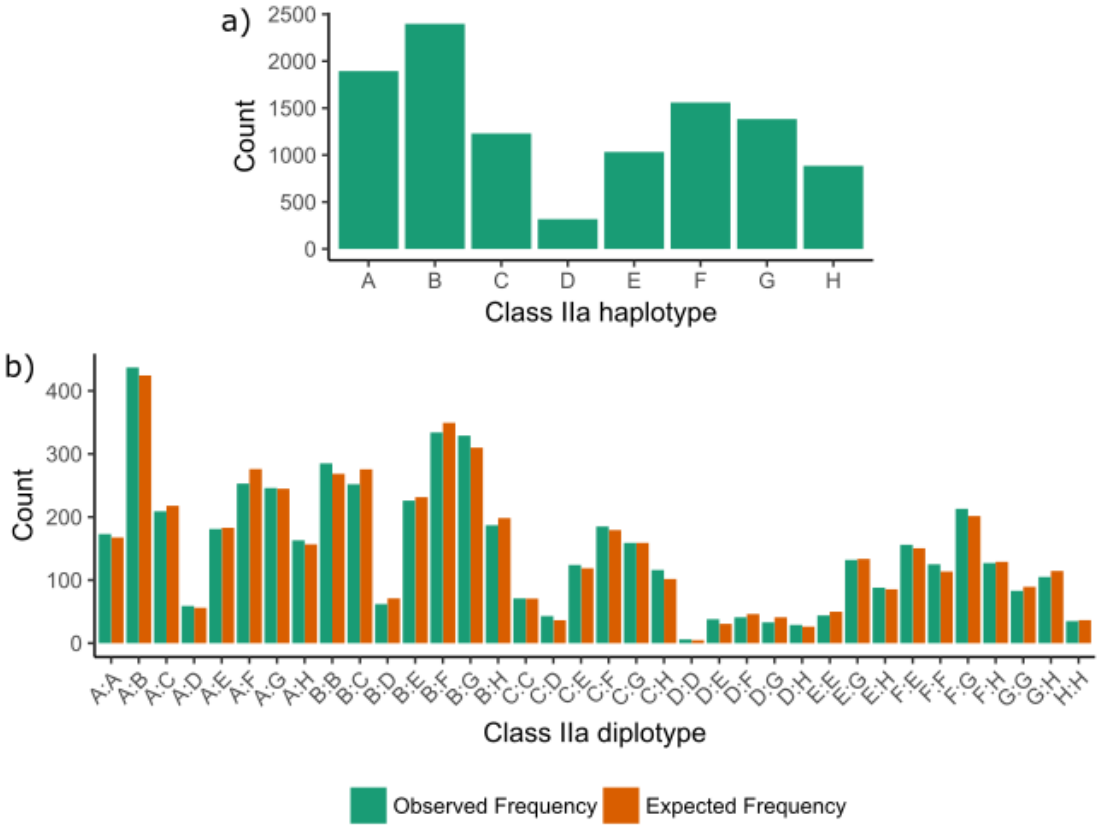



Figure 4.5. Haplotype (a) and diplotype (b) observed and expected frequencies under HWE for the 5349 individuals remaining after quality control, removal of novel haplotypes, and recoding of recombinant haplotypes.

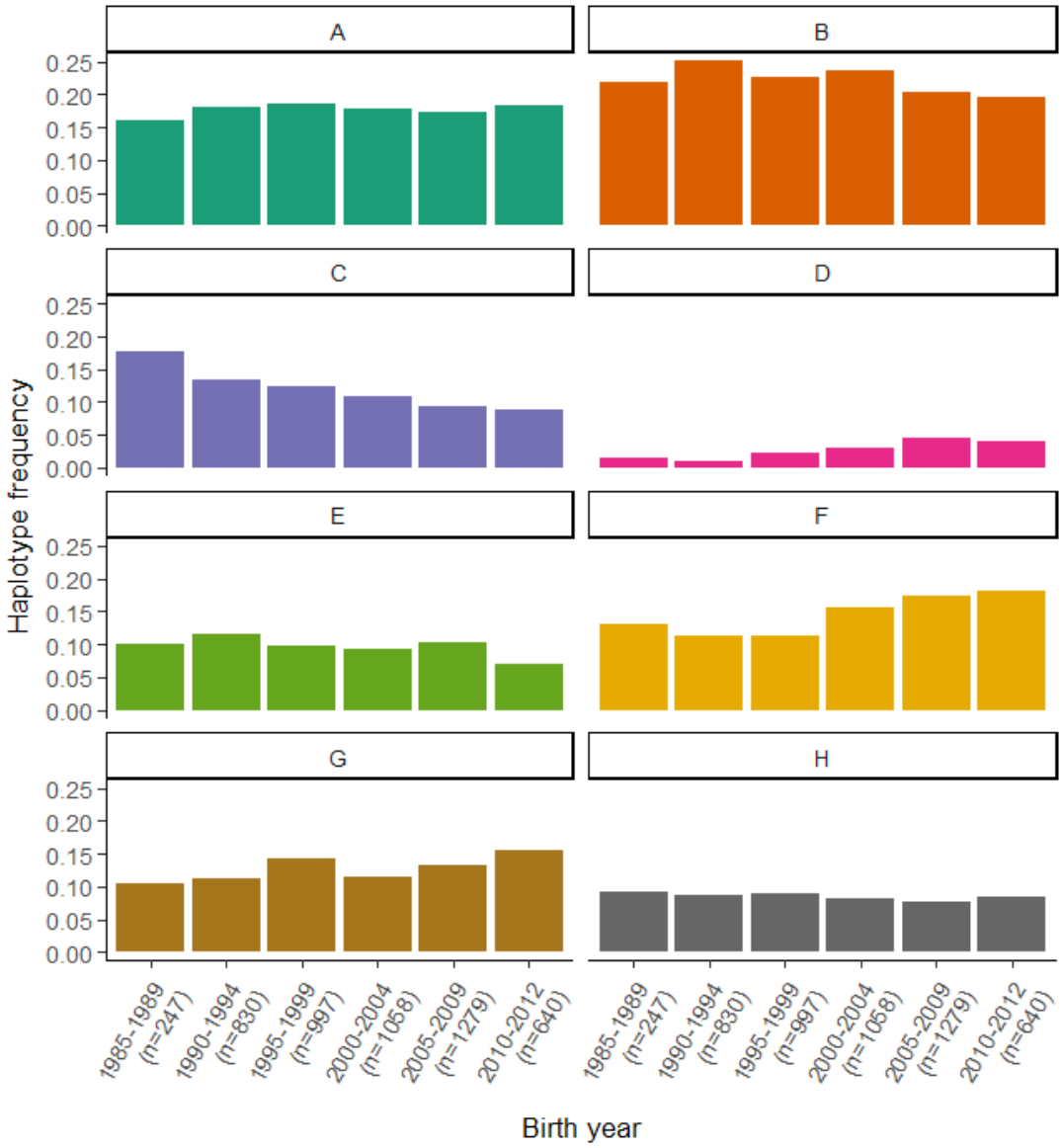



Figure 4.6. Haplotype frequencies in individuals born in each 5-year period. Sample sizes are shown below the birth year category, homozygous diplotypes were only counted once. Note that the final category includes only three cohorts as individuals born after 2012 have not been genotyped.

### 4.4.4 Ewens-Watterson tests and HWE

The number of MHC haplotyped individuals in each year was found to be a good approximation of the Village Bay population estimate for each year between 1989 and 2012 (except 2010 due to high genotyping failures for that cohort the DNA from which was largely contained the excluded plates 1-3; Figure 4.7).

Ewens-Watterson tests showed significantly negative $F_{nd}$ values for each LH stage (Table 4.4) and for the standing population in all years (Table 4.5). Significantly negative $F_{nd}$ values indicate balancing selection.

HWE was calculated across four LH stages and within the standing population of each year. For the four life history stages, the HWE tests were all non-significant for both the exact tests and heterozygote excess tests (Table 4.4). When considering each cohort separately, 'known conceived' for 2012 was marginally out of HWE for the exact test ($p$ = 0.023, α threshold = 0.025), with more homozygotes than expected ($H_E$ = 235.30, $H_O$ = 223), as was 'live born' for the heterozygote excess test for 2004 ($p$ = 0.021, α threshold = 0.025), also with more homozygotes than expected ($H_E$ = 232.02, $H_O$ = 241). All other years and tests for each LH did not deviate from HWE (Appendix E). The standing population was not found to deviate from HWE for any year in either test (Table 4.5).

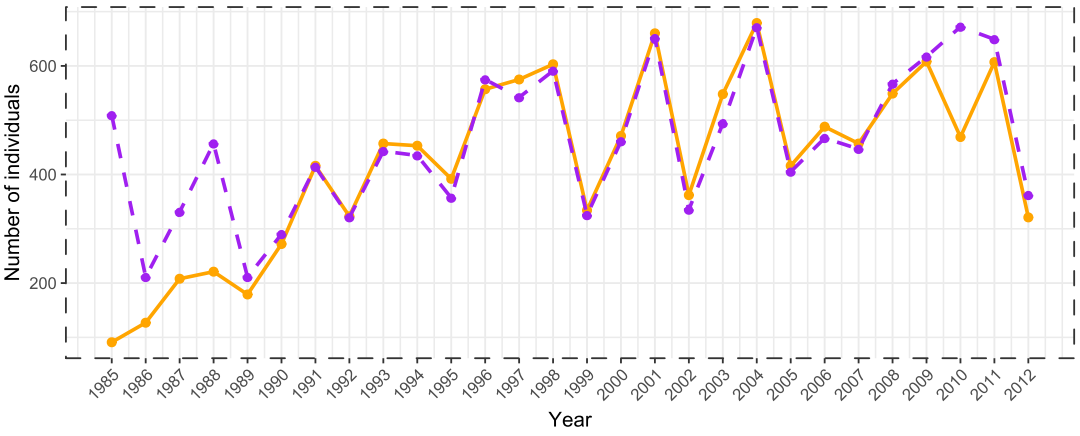



Figure 4.7. Sample sizes of MHC haplotype individuals in a given year (solid, orange line) compared to the estimated Village Bay population size of that year (dashed, purple line). Note that the Village Bay population estimates are calculated as individuals within Village Bay in October, whereas the standing population was determined here as individuals known to be alive in August.

Table 4.4. Results from Ewens-Watterson tests and HWE tests (exact and heterozygote excess – p values) for each LH stage. Sample sizes (number of individuals with MHC genotypes) are shown (n). Significant $p$ values are highlighted in bold.

| | | Ewens-Watterson tests | | | | HWE | |
|---|---|---|---|---|---|---|---|
| LH stage | n | $F_{obs}$ | $F_{exp}$ | $F_{nd}$ | $P$ value | Exact test | Heterozygote excess |
| Known conceived | 5158 | 0.1494 | 0.5475 | -2.0438 | < 0.0001 | 0.8275 | 0.7654 |
| Live born | 4643 | 0.1499 | 0.5443 | -2.0335 | < 0.0001 | 0.7738 | 0.7462 |
| Survived to 1st August | 3872 | 0.1493 | 0.5385 | -2.0220 | < 0.0001 | 0.6856 | 0.7193 |
| Survived to 2nd August | 1824 | 0.1480 | 0.5131 | -1.9681 | < 0.0001 | 0.8801 | 0.4257 |

Table 4.5. Results from HWE tests (exact and heterozygote excess – p values) and Ewens-Watterson tests for the standing population in any given year. Sample sizes (number of individuals with MHC genotypes) are shown (n). Significant *p* values are highlighted in bold.

| | | Ewens-Watterson tests | | | | HWE | |
|---|---|---|---|---|---|---|---|
| Year | n | $F_{obs}$ | $F_{exp}$ | *F*nd | *P* value | Exact test | Heterozygote excess |
| 1985 | 91 | 0.1595 | 0.4172 | -1.7359 | 0.0002 | 0.6242 | 0.1347 |
| 1986 | 127 | 0.1551 | 0.3932 | -1.6754 | 0.0003 | 0.8535 | 0.2719 |
| 1987 | 208 | 0.1523 | 0.4199 | -1.7466 | 0.0002 | 0.9252 | 0.3277 |
| 1988 | 221 | 0.1505 | 0.4230 | -1.7646 | 0.0001 | 0.7885 | 0.3254 |
| 1989 | 179 | 0.1546 | 0.4120 | -1.7161 | 0.0002 | 0.2399 | 0.6808 |
| 1990 | 272 | 0.1532 | 0.4334 | -1.7679 | 0.0002 | 0.6720 | 0.3951 |
| 1991 | 416 | 0.1594 | 0.4535 | -1.7717 | 0.0002 | 0.8801 | 0.2036 |
| 1992 | 323 | 0.1552 | 0.4417 | -1.7728 | 0.0001 | 0.7372 | 0.1408 |
| 1993 | 457 | 0.1595 | 0.4578 | -1.7805 | 0.0001 | 0.9664 | 0.1883 |
| 1994 | 453 | 0.1582 | 0.4574 | -1.7874 | 0.0001 | 0.9195 | 0.2702 |
| 1995 | 392 | 0.1544 | 0.4508 | -1.7968 | 0.0001 | 0.7951 | 0.0218 |
| 1996 | 557 | 0.1536 | 0.4665 | -1.8343 | < 0.0001 | 0.8551 | 0.2383 |
| 1997 | 575 | 0.1523 | 0.4678 | -1.8444 | < 0.0001 | 0.2523 | 0.0211 |
| 1998 | 603 | 0.1523 | 0.4699 | -1.8482 | < 0.0001 | 0.5236 | 0.5764 |
| 1999 | 334 | 0.1514 | 0.4433 | -1.7994 | 0.0001 | 0.6711 | 0.6157 |
| 2000 | 471 | 0.1545 | 0.4591 | -1.8130 | 0.0001 | 0.8828 | 0.6200 |
| 2001 | 660 | 0.1551 | 0.4737 | -1.8410 | < 0.0001 | 0.9047 | 0.6852 |
| 2002 | 362 | 0.1497 | 0.4471 | -1.8173 | 0.0001 | 0.9435 | 0.8358 |
| 2003 | 548 | 0.1487 | 0.4658 | -1.8614 | < 0.0001 | 0.5603 | 0.9669 |
| 2004 | 679 | 0.1507 | 0.4749 | -1.8688 | < 0.0001 | 0.1087 | 0.6413 |
| 2005 | 416 | 0.1470 | 0.4535 | -1.8464 | < 0.0001 | 0.1848 | 0.9490 |
| 2006 | 488 | 0.1457 | 0.4607 | -1.8686 | < 0.0001 | 0.5427 | 0.7053 |
| 2007 | 457 | 0.1456 | 0.4578 | -1.8632 | < 0.0001 | 0.9135 | 0.6115 |
| 2008 | 549 | 0.1439 | 0.4658 | -1.8896 | < 0.0001 | 0.2914 | 0.6082 |
| 2009 | 608 | 0.1477 | 0.4702 | -1.8767 | < 0.0001 | 0.9401 | 0.5840 |
| 2010 | 469 | 0.1510 | 0.4589 | -1.8337 | < 0.0001 | 0.8062 | 0.5108 |
| 2011 | 607 | 0.1463 | 0.4702 | -1.8845 | < 0.0001 | 0.4780 | 0.5176 |
| 2012 | 321 | 0.1509 | 0.4414 | -1.7983 | 0.0001 | 0.4144 | 0.6912 |

## 4.5 Discussion

This chapter describes the generation of, to the best of my knowledge, the largest non-human set of MHC data for a single population. Of 5951 individuals sampled within Village Bay between 1985 and 2012, MHC diplotype inference was successful for 5349 individuals, representing a success rate of 89.2 %. There were 276 individuals excluded because they were only genotyped on plates 1-3, an additional 297 due to low genotyping success (> 3/11 missing MHC SNP genotypes), 26 due to the inference of

novel haplotypes, and 3 due to mismatches with parental haplotypes. Although 10.8 % of the sampled individuals failed to be haplotyped, for those which passed the QC measures described here, haplotypes can be regarded as high quality and highly accurate, which will be beneficial to any subsequent population genetic and statistical analyses.

### 4.5.1 Genotyping errors

The repetition of a single individual on almost all DNA plates was valuable in detecting systematic failures and assessing the genotyping error rate. The high error rate (9 out of 22 loci) for the golden sheep on plate 3 supported the conclusion that there were systematic errors across plates 1-3. The genotyping error rate for the golden sheep across the remaining 57 plates on which it was successfully typed was only 0.48 %, suggesting that the error rate across individuals typed only once is likely to be similarly low. Whilst it is important for genotyping to be 100 % accurate for some applications, for instance forensic analyses, a large population study such as this would not be feasible under such a requirement and a compromise must be made between time, cost and accuracy.

The Mendelian inheritance check demonstrated that the phased haplotypes were consistent between parents and offspring, with only three exceptions. For one individual, 4597, SNP-level genotype inconsistencies with parents were identified at three loci, and at a single locus for individual 5864, although only a single parent was known. Whilst this suggests that genotyping errors can lead to incorrect haplotype inferences, in this population, with good pedigree information, it is clearly possible to detect such errors.

### 4.5.2 Novel haplotypes

There were 20 novel haplotypes identified in 26 individuals (0.5 % of haplotyped individuals), which were probably due to genotyping errors. Eight of these individuals were identified as having genotype Mendelian inheritance mismatches to at least one parent for at least one SNP locus. Some of these mismatches could have been caused by allelic dropout. At the haplotype level, all of the individuals carrying novel haplotypes and at least one genotyped parent or offspring were identified as having Mendelian inheritance mismatches. In all cases, the novel haplotype profiles were only a single base different from a known SNP profile (Table 4.2), and so the most parsimonious

explanation would be a genotyping error. The positions of these base changes were not consistent, suggesting that no one SNP locus was performing particularly poorly. Six individuals carrying novel haplotypes had between one and three SNP genotypes missing, and so poor imputation could also have generated these errors, although the overall error rate appears to be low and detectable using the QC methods carried out here.

### 4.5.3 Origin of the recombination event

The recombination event between haplotypes B and H identified in chapter 3 was found in an additional 55 related individuals, and could be traced back through the pedigree to a gametogenesis event in the mother of individual 4179. This recombinant SNP profile was identified initially in individual 4179 because he was genotyped on the Ovine Infinium HD Beadchip. Individuals were selected for genotyping on this chip because they were highly representative of the genetic variation within the Soay sheep population (Johnston *et al.* 2016). Individual 4179 was very prolific with a large number of descendants, hence his inclusion on the HD SNP chip. Of 4179's 53 known offspring, 45 have been haplotyped, and 26 were identified as carrying the recombinant haplotype. An additional two offspring of 4179 were imputed as carrying the recombinant haplotype as their own offspring carried it (see Figure 4.4). This highlights the value in having used individuals that were selected to be highly representative of the variation in the population to select the SNPs for the MHC panel. Were this recombination event not identified, and therefore not detectable, a substantial number of haplotypes would have been incorrectly imputed, which could have impacted subsequent analyses.

### 4.5.4 Haplotype and diplotype frequencies

All possible diplotypes were identified within the population, although the rarest, homozygote for D, was only identified in six out of the 5349 individuals genotyped (0.11 %). Over the study period (1985 – 2012), 5137 haplotyped individuals had known birth years (or expected birth years in the case of foetuses). There was no evidence of any dramatic shifts in haplotype frequencies of these individuals, which include the haplotypes of all individuals known to have been conceived. Figure 4.6 shows there may be a slight decrease in haplotype C and a slight increase in haplotype D which might warrant further investigation to identify whether these shifts are significant, and if so, if they are due to natural or selective forces.

### 4.5.5 Ewens-Watterson tests and HWE

The Ewens-Watterson tests were significantly negative for all LH stages, and within every year (1985-2012) for the standing population, and remained significant after sequential Bonferroni correction. This means that the observed level of homozygosity was consistently less than the expected homozygosity under neutrality, which indicates that allele frequencies were more even than expected, due to balancing selection (Ewens 1972; Watterson 1978). This pattern of lack of deviation from HWE but evidence of balancing selection using the Ewens-Watterson test was also observed at the *DRB1* microsatellite OLADRB (Paterson 1998). The Ewens-Watterson test assumes a constant population size (Spurgin & Richardson 2010), but the studied Soay sheep population on the island of Hirta was founded from 85 breeding individuals in 1932 (Boyd 1953; Boyd *et al.* 1964), with population growth interspersed with high mortality years since (Clutton-Brock *et al.* 2004a). It is unlikely that these demographic processes could have caused the observed pattern of even haplotype frequencies at the MHC class II haplotypes identified here as similar patterns would be expected at neutral loci, and no such evidence was found by Ewens-Watterson tests in MHC flanking microsatellites (Paterson 1998) and neutral microsatellites on chromosomes other than 20 where the MHC is located (Charbonnel & Pemberton 2005).

There was no evidence of deviations from HWE at any of the LH stages tested here (Table 4.4), nor within the standing population of any given year (Table 4.5). Whilst HWE was also tested for in each cohort at each LH stage (Appendix E), sample sizes were often small, especially in the 'survived to second August' category, which may limit the power to detect any deviations. Deviations from HWE were identified in the exact test for the 'known conceived' stage in 2012 and for the heterozygote excess test in at the 'live born' stage in 2004 (both an excess of homozygotes); however, significance was marginal after applying sequential Bonferroni correction and so some caution should be applied in considering this evidence of deviation from HWE. Whilst deviation from HWE can indicate selective or demographic processes operating within the population (Guo & Thompson 1992; Spurgin & Richardson 2010), lack of deviation from HWE does not necessarily provide good evidence to the contrary (Spurgin & Richardson 2010).

# 5. Associations between haplotypes and phenotypic traits associated with health

## 5.1 Abstract

Pathogen-mediated selection (PMS) is thought to maintain the high levels of diversity observed at the MHC in many species. There are three key modes of selection thought to operate under PMS: heterozygote advantage, negative frequency-dependent selection and fluctuating selection. Testing these hypotheses requires locus-specific genotypes across multiple loci for large numbers of individuals, as well as corresponding phenotype data. In previous chapters, the class IIa haplotypes were characterised in the Soay sheep, and diplotypes were generated for over 5000 individuals. Because the Soay sheep of St. Kilda have been intensively monitored at an individual level since 1985, there is also a wealth of corresponding samples and data, such as weight, immune measures and parasitological data. Mixed effects models were used to investigate the relationships between heterozygosity or haplotypes and three ways of measuring health status (weight, anti-*Teladorsagia circumcincta* IgA, IgE and IgG, and faecal egg count – FEC), whilst accounting for potentially confounding effects. Analyses were also carried out within an animal model framework to incorporate a relatedness matrix and account for additive genetic variance for the traits. Heterozygosity was not found to be associated to any of the five traits analysed. Numerous associations were identified between the presence of a haplotype and all five traits in different age-sex classes. By carrying out analyses using both standard generalised linear mixed models and animal models, it was clear that a number of apparently strong associations were lost after accounting for relatedness. However, from a highly conservative perspective, where an association is considered significant at the Bonferroni threshold within an animal model, five haplotypes (B, C, E, F and G) were associated with variation in August weight, IgE titres, IgG titres and FEC in yearlings and adults, but no strong associations were identified in lambs. The lack of associations between heterozygosity and traits suggests that heterozygote advantage is unlikely to be operating. Associations between specific haplotypes and traits, on the other hand, suggests that negative frequency-dependent or fluctuating selection is likely operating in the Soay sheep.

## 5.2 Introduction

Diversity at MHC genes is remarkably high. Substantial effort has been made by many researchers to investigate how such high levels of variation are maintained, across a wide variety of vertebrate taxa in both natural and experimental conditions, and through theoretical modelling (as discussed in Chapter 1, Appendix A). Pathogen-mediated selection (PMS) is frequently implicated in such studies (Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). The key PMS mechanisms involved, which are not necessarily mutually exclusive, are heterozygote advantage (Doherty & Zinkernagel 1975; Slade & McCallum 1992) and variation in selection pressure through either negative frequency-dependent selection (NFDS; (Apanius *et al.* 1997)) or fluctuating selection (Hill 1991; Hedrick 2002). Heterozygote advantage occurs when heterozygotes have greater fitness than either homozygote class (Takahata & Nei 1990; Hughes & Nei 1992; McClelland, Penn & Potts 2003). NFDS occurs due to rare allele advantage. Pathogens are predicted to be under selection pressure to evade the most common MHC genotypes, resulting in rare MHC genotypes enjoying a selective advantage, and creating a cyclical co-evolutionary arms race (Takahata & Nei 1990; Slade & McCallum 1992). Under fluctuating selection (FS), directional selection due to variation in pathogen pressure varies in time and space such that it maintains diversity (Hedrick 2002).

Testing these hypotheses is challenging, both experimentally and within wild systems. Experimental setups are rarely capable of replicating the wide array of pathogens and parasites that occurs within a wild host (although see Kalbe *et al.* 2009; Eizaguirre *et al.* 2012a), and are often therefore limited in the conclusions that they can draw about natural processes. Conversely, in wild systems, MHC diversity combined with variation in genetic background, pathogen diversity and the different environmental conditions experienced by different individuals over their ontologies makes it difficult for such studies to disentangle the effects of MHC variation and the selection pressures acting upon it.

Evidence of selection on MHC genes can be investigated by looking for the signature of historical balancing selection or by looking for contemporary selection across multiple populations or within a single population (Piertney & Oliver 2006; Spurgin & Richardson 2010). Nucleotide variation in expressed regions can reveal information about the history of selection on MHC alleles. An increased ratio of non-synonymous to

synonymous substitutions (dN/dS) is an indicator of positive selection in a DNA sequence (Hill & Hastie 1987; Hughes & Nei 1992). Patterns of MHC diversity amongst different populations compared with diversity in genome regions deemed to be neutral can reveal information about contemporary selection, for example more even allele frequencies at MHC than neutral markers within a population may suggest similar selection pressures at the MHC, whilst greater divergence at MHC than at neutral markers between populations might indicate adaptive variation and therefore divergent selection. Both patterns, however, suggest balancing selection within the species, likely NFDS or FS acting at either the species level, in the first instance, or the population level (Spurgin & Richardson 2010). However, to investigate if contemporary selection is acting within a single population, it is necessary to assess the temporal or spatial variation in MHC allele frequencies (NFDS or FS), look for directional selection on different alleles or haplotypes (NFDS or FS) or to look for associations between MHC genotypes and phenotypes (NFDS or FS) or heterozygosity. Separating the processes of NFDS and FS, however, is more challenging as the effects on patterns of MHC diversity are very similar, may require observations of both pathogens and hosts over evolutionary time, and multiple mechanisms may be acting on a population at the same time (Spurgin & Richardson 2010).

This study investigates the associations between MHC class IIa haplotypes and phenotypes related to health in an unmanaged population of Soay sheep on St. Kilda. The Soay sheep are descendants of an early domestication of sheep (*Ovis aries*) but have lived on the islands of the St Kilda archipelago virtually unmanaged, and therefore subject to natural and sexual selection for several thousand years (Clutton-Brock *et al.* 2004b). The population does not have any predators or competitors for food, and undergoes regular density-dependent population crashes, when mortality is related to winter weather, density, age, sex (Coulson *et al.* 2001) and parasite burden (Clutton-Brock *et al.* 2004a). Gastrointestinal nematodes (GIN) are prevalent in the Soay sheep, with virtually 100 % prevalence in lambs, and immunity to GIN develops with age, which is more pronounced in females than males (Wilson *et al.* 2004; Craig, Pilkington & Pemberton 2006; Craig *et al.* 2008). GIN are a major selective force on the Soay sheep (Gulland 1992; Craig *et al.* 2006; Hayward *et al.* 2011), and parasite burden, measured as faecal egg counts (FEC), is negatively associated with body weight and size (Coltman *et al.* 2001a; Craig *et al.* 2008) and over-winter survival (Gulland 1992; Gulland & Fox 1992; Gulland *et al.* 1993; Illius *et al.* 1995; Hayward *et al.* 2011). FEC is heritable in

lambs (0.26 ± 0.12 (Beraldi *et al.* 2007)) and lowly heritable in adults (Females: 0.0 – 0.05 ± 0.02, Males: 0.0 – 0.12 ± 0.03 (Beraldi *et al.* 2007; Robinson *et al.* 2009)), demonstrating that there is a genetic basis to GIN resistance.

In Soay sheep, the relationships between MHC class IIa variation and FEC or over-winter survival in lambs and yearlings have been investigated previously using the microsatellite OLADRB located within the second intron of the class II MHC *DRB1* gene (Paterson 1998; Charbonnel & Pemberton 2005). Paterson *et al.* (1998) showed that, between 1985 and 1994, the OLADRB 257 bp allele was associated with both increased FEC and reduced survival in lambs, whilst the OLADRB 263 bp allele was associated with both decreased FEC and increased survival in yearlings. Overall heterozygosity at OLADRB was not found to be associated with either trait. Charbonnel and Pemberton (2005) showed that there was some evidence of spatial differentiation in the OLADRB microsatellite between three spatially defined areas of the population which was not evident at five presumably neutral loci, and evidence of temporal variation in one of the three areas, which may indicate fluctuating selection is operating. While OLADRB does reflect functional variation at *DRB1* (Paterson 1998), the characterisation of multiple class IIa loci (Chapter 2 and see Appendix C) shows that the *DRB1* locus or OLADRB alone are not fully predictive of class IIa haplotypic variation.

This study aims to re-investigate the relationships between MHC class IIa variation and FEC in Soay sheep, as well as with other phenotypic measures associated with health, using the haplotype data generated for thousands of individuals via KASP (Chapter 4). This includes an additional 18 years of data, compared to the Paterson *et al.* (1998) study which greatly increases the sample size of lambs and yearlings with matched genotypic and phenotypic data. Additionally, this work increases samples sizes of adults to enable association studies to be carried out in this age class. MHC studies have rarely been carried out in adult sheep (but see Martinez-Valladares *et al.* 2005), probably due to a focus on lamb productivity, as well as reduced GIN burden in adults compared to lambs (Craig *et al.* 2008).

In this chapter, associations between MHC class IIa haplotypes and three ways of measuring health status (five traits in all) in Soay sheep. August weight is negatively correlated with FEC (Coltman *et al.* 2001a; Craig *et al.* 2008; Hayward *et al.* 2014b) and positively correlated with fitness (Milner *et al.* 1999; Jones *et al.* 2005). Anti-*Teladorsagia circumcincta* (Anti-*T. circ*) immunoglobulin (Ig) A is negatively associated with FEC (Coltman *et al.* 2001b), anti-*T. circ* IgE is positively associated with FEC

(Nussey *et al.* 2014), and anti-*T. circ* IgG is positively associated with survival (Nussey *et al.* 2014; Watson 2016). FEC is negatively correlated with survival in lambs (Hayward *et al.* 2011). Associations between haplotypes and traits should, therefore, be indicative of selection.

## 5.3 Methods

### 5.3.1 General field data collection

The Village Bay population of Soay sheep has been intensively studied since 1985 (see Chapter 1). Lambing (April and May) is monitored to determine day of birth, litter size and mother's ID, and approximately 90 % of lambs are caught shortly after birth, tagged for individual identification, and phenotypic measurements, blood and ear punch samples are collected. In August, 50 – 60 % of the population is rounded up and phenotypic measurements (including body weight), blood and faecal samples are collected. Small samples of individuals which were treated with anti-helminthic boluses or drenches for experimental purposes (1991-1992, 1995 and 2001-2002; (Wilson *et al.* 2004; Craig *et al.* 2009)) were excluded from analyses for all years following treatment.

### 5.3.2 Pedigree

Maternal ID is primarily determined by observation during lambing. Where this has not been possible, for example lambs born after the lambing team has left, maternal ID is assigned by observation of mother and lamb during the August catch, when lambs are still closely associated with and suckled by their mothers. If an observational maternal ID cannot be obtained, for example for neonates found dead on the hill, it is determined genetically. Paternal ID is always determined genetically. Genetic parentage assignments were carried out by Bérénos *et al.* (2014). Briefly, assignments were inferred from 315 unlinked SNPs within MASTERBAYES (Hadfield, Richardson & Burke 2006). Additionally, the R package SEQUOIA (Huisman 2017) was used to cluster half-siblings which share a parent that has not been genotyped, by assigning dummy parent IDs. The combined observational and genetic pedigree includes 7447 individuals (including 332 dummy IDs), 7014 paternal links (including 722 dummy ID links) and 6229 maternal links.

### 5.3.3 Data analyses

MHC class IIa haplotypes were tested for associations with variation in five health-related traits: August weight, three different plasma immunoglobulin measures and strongyle parasite burden, using general or generalised linear mixed effects models (GLMM). This enabled any effect of the haplotypes to be assessed after accounting for other variables known to affect the trait of interest by including them in the model. The fixed and random effects considered for each trait are detailed below.

Initially, a non-genetic model was determined for each trait by assessing the significance of the fixed effects by deletion testing, where the full model is compared to a reduced model lacking the term of interest, and comparing the likelihood ratios of the two models. The terms tested are shown in Table 5.1. The least significant term was dropped and terms were re-tested until all terms remained significant. Fixed effects of sex and a measure of genome-wide inbreeding Fhat3, (Yang *et al.* 2011), which is strongly correlated with genome-wide heterozygosity and has been shown depress several traits including fitness components (Berenos et al 2016), were retained in the model even if they were not significant to ensure that any MHC associations identified were over and above any effects of sex or genome-wide inbreeding. The significance of random effects was tested in the same manner, but they were only excluded from a model if the variance explained by the term was extremely small and not-significant and they were thought unlikely to be important based on first principles.

Following selection of the non-genetic model, the associations between the MHC class II haplotypes and traits were tested by adding the term as a fixed effect and deletion testing as for the non-genetic models. Three different tests were performed for each response variable. Firstly, to test whether heterozygosity was associated with trait variation, heterozygosity was fitted as a two-level factor (heterozygous or homozygous, irrespective of the diplotype). Secondly, to determine whether particular haplotypes were associated with traits, two separate models were constructed for each haplotype. These two models were: i) presence/absence of a haplotype, fitted as a two-level factor where each individual either carried the haplotype (1) or did not (0) and ii) as dosage of a particular haplotype, where sample size permitted (≥ 10 homozygotes within an age or age-sex class), fitted as a three-level factor where an individual either does not carry the haplotype (0), is heterozygous for the haplotype irrespective of the partner haplotype (1) or homozygous for the haplotype (2). For each of the haplotype-specific tests, a Bonferroni correction was applied (as $\alpha = 0.05/8 = 0.00625$, since there were 8

haplotypes) to reduce the likelihood of type I errors arising due to multiple testing. Thus, an association between a haplotype and trait was significant if $p < 0.0065$, and was referred to as marginal or a trend where $0.00625 < p < 0.05$. Because males and females may invest differently in their immune systems and MHC alleles may therefore be under different levels of selection in the two sexes, I tested for haplotype by sex interactions where separate-sex models were not already being constructed. Very few of haplotype-by-sex interactions were significant and none after Bonferroni correction; the significance test for each interaction model is included in Appendix F but results are not shown here.

Because it is an individual-based project studying all individuals living in a defined study area, the Soay sheep data set includes related individuals and this has potentially important ramifications for study selection on specific genome regions. Female Soay sheep are philopatric so the female population consists of a range of female relatives. The mating system is highly promiscuous so that lambs born in the same or adjacent cohorts may share the same sire. Therefore, there is a risk that estimates of MHC haplotype effects will be biased by including related individuals in the GLMM analysis – for example a rare haplotype may occur at high frequency in a lineage which, due to shared genetics elsewhere in the genome, has extreme breeding values for a trait. To account for this, the putative haplotype effects were also assessed within an animal model (AM) framework, in which a relatedness matrix based on the pedigree is fitted as a random effect and the additive genetic variance for a trait (and hence its heritability) can be estimated. This approach enabled separation of any haplotype-trait associations from additive genetic variance for the traits due to effects elsewhere in the genome.

All statistical analyses were carried out in R version 3.4.0 (R Core Team 2017). General linear mixed effects models were carried out in the package lme 4 version 1.1-13 (Bates *et al.* 2015) and generalised linear mixed effects models in glmmADMB version 0.8.4 (Fournier *et al.* 2012). Animal models were carried out using ASReml-R version 3.0 (Gilmore *et al.* 2009).

Table 5.1. Fixed and random effects tested in mixed effects models of health-related traits. Variables highlighted in grey were included as fixed effects and those

| | | August body weight | | | Immunoglobulins | | FEC | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Lambs | Yearlings | Adults | Lambs | Adults | Lambs | Yearlings | Adults |
| Fixed effects | Sex | Y | Y | Y | Y | Y | Y | Y | Y |
| | Litter size | Y | | | | | Y | | |
| | Age[a] | Y (Days) | | Y (Years) | Y (Days) | Y (Years) | Y (Days) | | Y (Years) |
| | Age$^2$ | | | Y | | Y | | | Y |
| | Fhat3 | Y | Y | Y | | | Y | Y | Y |
| | Birth weight | Y | | | | | Y | | |
| | August weight[b] | | | | | | Y | Y | Y |
| | Maternal Age | | | | | | Y | | |
| | Maternal Age$^2$ | | | | | | Y | | |
| Random effects | ID | | | Y | | Y | | | Y |
| | Birth year | Y | Y | Y | Y | Y | Y | | Y |
| | Capture year | | | Y | | Y | | Y | Y |
| | Maternal ID | Y | | | Y | | Y | Y | Y |
| | ELISA plate ID | | | | Y | Y | | | |
| | ELISA plate Run Date | | | | Y | Y | | | |

[a] Mean centred for lamb models
[b] Mean centred

## 5.3.4 The trait data

### 5.3.4.1 August body weight

During August, all sheep that are caught are weighed to the nearest 0.1 kg. August weight was normally distributed and modelled with a Gaussian error structure. Heritability of August weight is known to vary with age (Wilson *et al.* 2007). Separate models were run for the different age classes (lambs, yearlings and adults) (Bérénos *et al.* 2014). Distribution of August weight measures in each age and sex class are shown in Figure 5.1 and sample numbers of individuals in each age and haplotype class are shown in Table 5.2.

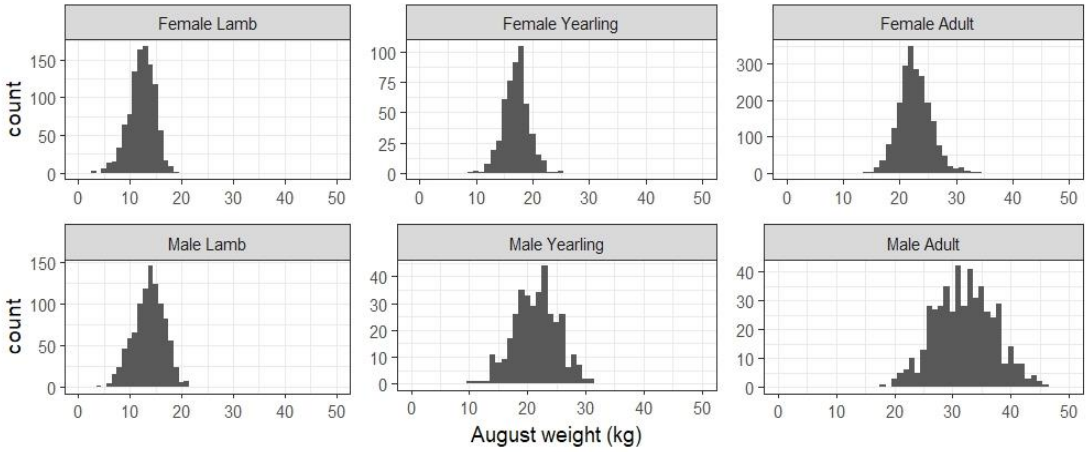



Figure 5.1. Histograms showing the distribution of August weight (kg) measures in each age and sex class.

Table 5.2. Number of individuals within each haplotype-dosage category. For presence/absence-of-haplotype models, present category (1) is the sum of $n_1$ and $n_2$.

| | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Age | Dosage category | A | B | C | D | E | F | G | H |
| Lambs | $n_0$ – absent | 1047 | 924 | 1185 | 1398 | 1196 | 1078 | 1117 | 1268 |
| | $n_1$ – 1 copy | 404 | 498 | 300 | 96 | 291 | 384 | 354 | 219 |
| | $n_2$ – 2 copies | 47 | 76 | 13 | 4 | 11 | 36 | 27 | 11 |
| Yearlings | $n_0$ – absent | 598 | 531 | 672 | 812 | 690 | 640 | 650 | 717 |
| | $n_1$ – 1 copy | 248 | 298 | 183 | 54 | 168 | 209 | 202 | 142 |
| | $n_2$ – 2 copies | 20 | 37 | 11 | 0 | 8 | 17 | 14 | 7 |
| Adults | $n_0$ – absent | 628 | 579 | 698 | 871 | 747 | 682 | 705 | 795 |
| | $n_1$ – 1 copy | 277 | 300 | 214 | 56 | 173 | 227 | 210 | 125 |
| | $n_2$ – 2 copies | 23 | 49 | 16 | 1 | 8 | 19 | 13 | 8 |

### 5.3.4.2 Anti-*Teladorsagia circumcincta* immunoglobulin (Ig) measures

Immunoglobulin isotypes IgA, IgE and IgG are antibodies involved in the acquired immune response to GIN in sheep (Stear, Strain & Bishop 1999a; Lee *et al.* 2011; Hayward 2013). Parasite-specific IgA acts at mucosal surfaces and is known reduce worm growth and fecundity (Stear *et al.* 1999a, 2004; Gutiérrez-Gil *et al.* 2010; Lee *et al.* 2011). Parasite-specific IgE is also predominantly at mucosal surfaces and is involved in the degranulation of mast cells, which are white blood cells involved in parasite expulsion (Murphy *et al.* 2010). IgG is the primary plasma antibody which can interact directly with the parasite on ingestion of blood or gastric mucus.

During August captures, up to 9 mL of whole blood is collected into lithium heparin tubes and stored at 4 °C until processed, which occurs within 48 hours. Samples are centrifuged at 3000 r.p.m. (approximately 1500 *g*) for 10 mins, and the plasma is removed and stored at –20 °C. The activity of IgA, IgE and IgG against *Teladorsagia circumcincta* third stage larvae (L3) was measured by Kathryn Watt, Alex Sparks and Rona Sinclair using direct (IgA and IgG) and indirect (IgE) ELISAs, using a protocol similar to that described previously by Nussey *et al.* (2014), and described in full here.

The capture antigen for both the direct and indirect assays was *T. circumcincta* L3 somatic antigen (Moredun Research Institute). The capture antigen was diluted to 2 µL/mL in 0.06M carbonate buffer (pH 9.6), and 50 µL added to each well of a Nunc-immuno 96-microwell plate. The plate was covered and incubated overnight at 4 °C, and then washed with tris-buffered saline-Tween (TBST) using a plate washer. Soay sheep plasma sample was diluted 1:50 for IgA and IgE, and 1:12800 for IgG, and 50 µL added to each well. The plate was then covered and incubated at 37 °C for 1 hour, before washing five times with TBST.

For the direct ELISAs (IgA and IgG), 50 µL of 1:16000 diluted rabbit anti-sheep IgA detection antibody conjugated to horseradish peroxidase (HRP; AbD Serotec AHP949P) was added to each well, or 50 µL of 1:16000 diluted rabbit anti-sheep IgG detection antibody conjugated HRP (AbD Serotec 5184-2104) was added to each well. For the indirect IgE ELISA, 50 µL of 1:100 diluted anti-sheep IgE (mouse monoclonal IgG1, clone 2F1, provided by the Moredun Research Institute) was added to each well, incubated for 1 hour at 37 °C, washed five times with TBST, and then 50 µL of 1:8000 diluted (in TBST) goat anti-mouse IgG1-HRP detection antibody (AbD Serotec STAR132P) was added to each well.

All plates were then incubated for 1 hour at 37 °C, washed five times with TBST, and 100 µL of SureBlue TMB 1-Component microwell peroxidase substrate was added to each well and incubated in the dark for 5 minutes at 37 °C. Reactions were then stopped by adding 100 µL of 1M hydrochloric acid to each well. Optical densities (OD) were read immediately at 450 nm using a ThermoScientific GO Spectrophotometer.

A number of quality control measures were used throughout. Firstly, to minimise any confounding effects of capture year or age with variation between plates, each plate included samples from two randomly paired years, and each year included different age groups. Every plate was run in duplicate. Sample ODs from the duplicated plates

were compared, and if the variation was large, both samples were discarded. Additionally, if the correlation of paired ODs across the duplicated plates was $r < 0.8$, both plates were re-run. Secondly, every plate included two blank wells (containing 50 μL TBST instead of a plasma) and two positive controls – for the IgA and IgG assays these were plasma samples from healthy, non-immunised domestic sheep, and for the IgE assays they were serum from ewes trickle infected with *T. circumcincta*. The numbers of samples that failed quality control measures were 13 for IgA, 8 for IgE and 27 for IgG.

Due to the lack of standard solutions, all results were measured as OD values. The OD ratio of each sample was calculated as (sample OD – blank OD) / (positive control OD – blank OD). If the blank OD was greater than the sample OD, the OD ratio was set to zero to prevent negative OD ratios. The mean OD ratio was then calculated for each duplicated sample, and these values were used in all subsequent analyses.

The distribution of immunoglobulin measures differed between lambs and adults (see Figure 5.2). Lambs, which are aged 4 months when sampled, are expected to be still in the process of developing their acquired immune system, and therefore their immunoglobulin measures may be more likely to vary than older animals. As there was no difference in the distributions between yearlings and adults (data not shown), immunoglobulins were modelled separately for lambs and adults (aged 1 year and older). Distributions of anti-*T. circ* IgA, IgE and IgG were not normally distributed in lambs, although diagnostic plots showed that residuals were not too badly distributed for anti-*T. circ* IgA and IgG (not shown). Consequently, for lambs, analyses were carried out on both untransformed and Box-Cox transformed immunoglobulin measures. As transformation had almost no effect on coefficient estimates, results from models using untransformed responses are presented. Distributions of anti-*T. circ* IgA, IgE and IgG were normally distributed in adults and therefore not transformed. Distributions of measures for each age and sex category are shown for each immunoglobulin isotype in Figure 5.2, and numbers of individuals within each age sex and haplotype category in Table 5.2.

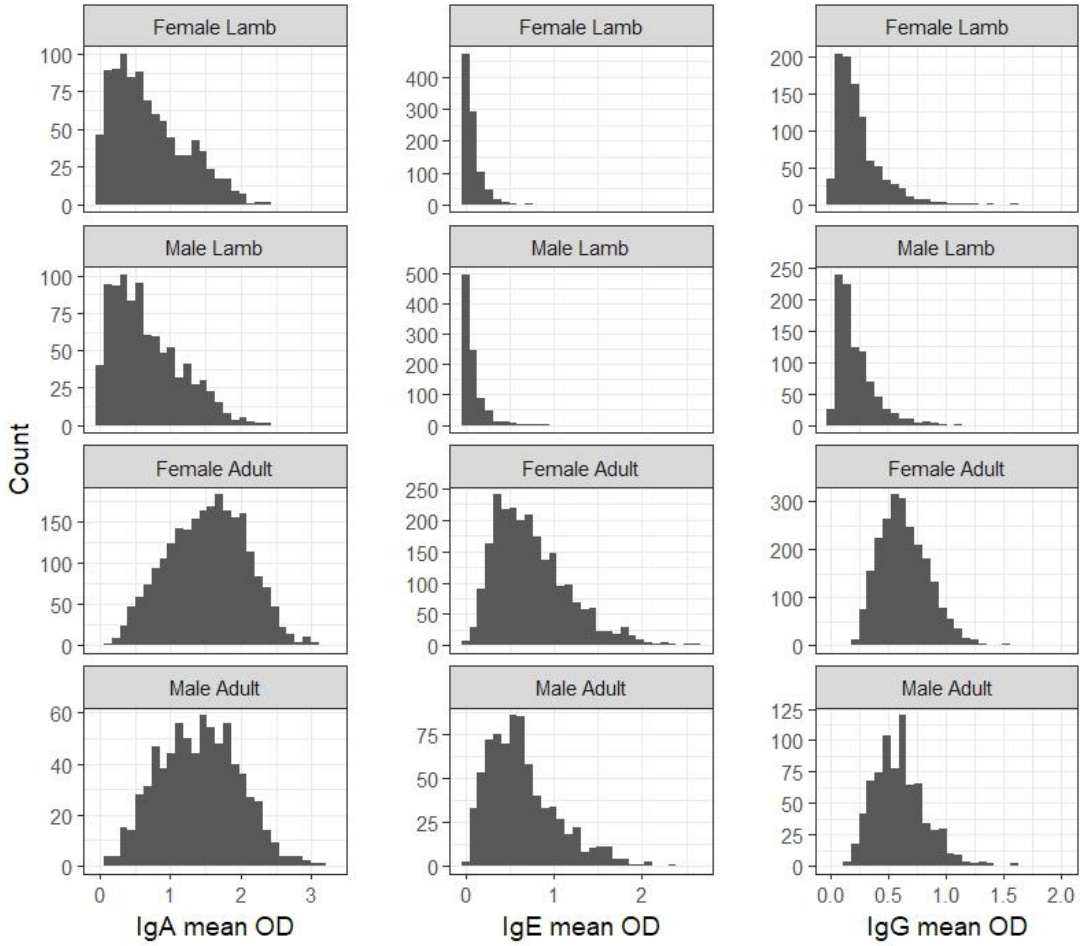



Figure 5.2. Distribution of mean OD for each immunoglobulin isotype in each age and sex category.

Table 5.3. Number of individuals within each haplotype-dosage category for immunoglobulin analyses. For presence/absence-of-haplotype models, present category (1) is the sum of $n_1$ and $n_2$.

| | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Age | Dosage category | A | B | C | D | E | F | G | H |
| Lambs | $n_0$ – absent | 1084 | 958 | 1256 | 1469 | 1260 | 1143 | 1177 | 1334 |
| | $n_1$ – 1 copy | 437 | 530 | 303 | 100 | 299 | 394 | 368 | 229 |
| | $n_2$ – 2 copies | 52 | 85 | 14 | 4 | 14 | 36 | 28 | 10 |
| Adults | $n_0$ – absent | 771 | 708 | 876 | 1078 | 917 | 845 | 871 | 956 |
| | $n_1$ – 1 copy | 337 | 379 | 249 | 63 | 214 | 275 | 253 | 174 |
| | $n_2$ – 2 copies | 34 | 55 | 17 | 1 | 11 | 22 | 18 | 12 |

#### 5.3.4.3 Faecal Egg Counts (FEC)

Faecal samples are collected from as many individuals as possible during capture and processing in August. Faecal egg counts (FEC) of nematode worms are performed using a modified McMaster technique (MAFF 1986), as described in Gulland & Fox (1992) and Wilson *et. al* (2004). This protocol enumerates Strongyle-type eggs per gram of wet weight faeces, but does not differentiate between several morphologically indistinguishable species. Of the species contributing to FEC on St Kilda, *Trichostrongylus axei, Trichostrongylus vitrinus* and *Teladorsagia circumcinta* eggs are the most abundant, but eggs from *Chabertia ovina, Bunostomum trigonocephalus*, as well as *Strongyloides papillosus* may also be present (Craig *et al.* 2006). This measure of FEC has been shown previously to be correlated with adult worm count within the abomasum in Soay sheep on St. Kilda and on Lundy (Grenfell *et al.* 1995; Boyd 1999; Wilson *et al.* 2004), and is therefore considered a good index of worm burden.

The distribution of FEC is known to differ between age classes and sexes in Soay sheep (Figure 5.3; (Gulland & Fox 1992; Wilson *et al.* 2004; Craig *et al.* 2006), and FEC declines with age in females but not males (Craig *et al.* 2006). In adults, FEC is repeatable and previous analyses have demonstrated heritable variation in FEC in Soay sheep (Coltman *et al.* 2001a; Beraldi *et al.* 2007; Hayward *et al.* 2014a; b). Separate models were run for lambs, yearlings and adults, and for each sex in yearlings and adults.

FEC is overdispersed in all sex-age groups, and was thus modelled with a negative binomial error structure within a generalised linear mixed effects model (Wilson & Grenfell 1997). In order to fit FEC within an animal model in ASREML-R, which cannot fit a negative binomial error structure, FEC was also log transformed, as Log(FEC+50), adding half the minimum detection limit (100 eggs/g). Model estimates from the mixed effects models (from here on termed GLMM) with Gaussian errors were compared to those from the GLMM with negative binomial errors, to show that estimates and confidence intervals were similar for the presence/absence-of-haplotype analyses but diverged for some haplotypes when analysing dosage (see Appendix F for details). Log(FEC+50) was then fitted within an AM in the same manner as for August weight and immunoglobulin models, but effects were interpreted with caution. Numbers of individuals within each sex, age and haplotype-dosage category are shown in Table 5.4.

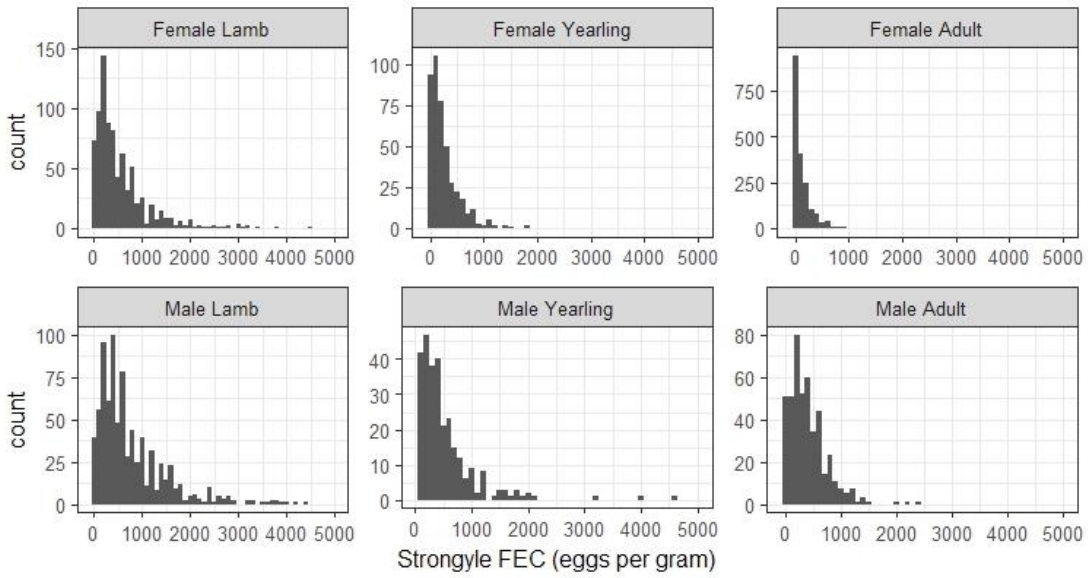



Figure 5.3. FEC distributions in each age and sex class.

Table 5.4. Number of individuals within each haplotype-dosage category for FEC analyses. For presence/absence-of-haplotype models, present category (1) is the sum of $n_1$ and $n_2$.

| Age | Sex | Dosage category | Haplotype A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|---|---|
| Lambs | Combined | $n_0$ – absent | 1111 | 968 | 1290 | 1529 | 1285 | 1197 | 1220 | 1379 |
| | | $n_1$ – 1 copy | 451 | 559 | 316 | 87 | 322 | 381 | 374 | 232 |
| | | $n_2$ – 2 copies | 58 | 93 | 14 | 4 | 13 | 42 | 26 | 9 |
| Yearlings | Females | $n_0$ – absent | 306 | 268 | 332 | 400 | 338 | 316 | 325 | 368 |
| | | $n_1$ – 1 copy | 119 | 144 | 96 | 33 | 90 | 109 | 102 | 63 |
| | | $n_2$ – 2 copies | 8 | 21 | 5 | 0 | 5 | 8 | 6 | 2 |
| | Males | $n_0$ – absent | 216 | 183 | 243 | 292 | 254 | 230 | 232 | 244 |
| | | $n_1$ – 1 copy | 82 | 112 | 61 | 16 | 53 | 70 | 71 | 59 |
| | | $n_2$ – 2 copies | 10 | 13 | 4 | 0 | 1 | 8 | 5 | 5 |
| Adults | Females | $n_0$ – absent | 416 | 390 | 464 | 581 | 497 | 464 | 471 | 533 |
| | | $n_1$ – 1 copy | 187 | 193 | 146 | 38 | 119 | 145 | 138 | 82 |
| | | $n_2$ – 2 copies | 17 | 37 | 10 | 1 | 4 | 11 | 11 | 5 |
| | Males | $n_0$ – absent | 159 | 146 | 186 | 218 | 193 | 163 | 179 | 199 |
| | | $n_1$ – 1 copy | 71 | 82 | 44 | 17 | 39 | 64 | 54 | 33 |
| | | $n_2$ – 2 copies | 5 | 7 | 5 | 0 | 3 | 8 | 2 | 3 |

# 5.4 Results

## 5.4.1 August weight

There was no association between MHC class IIa haplotype heterozygosity and August weight in lambs ($p$ = 0.658), yearlings ($p$ = 0.700), or adults ($p$ = 0.279). In lambs, there were also no associations between any of the MHC class IIa haplotypes and August weight (Figure 5.4a-b, Appendix F) when considering either presence or dosage of the haplotypes.

In yearlings Haplotypes E and G were associated with weight (Figure 5.4c-d, Appendix F). The presence of haplotype E was associated with reduced August weight (GLMM: $\beta_1$ = –0.756 ± 0.226, $p$ = 0.001; AM: $\beta_1$ = –0.738 ± 0.239, $p$ – 0.002, where $\beta_i$ ± SE, where $i$ denotes the presence/absence category or dosage category of the estimate), but dosage of haplotype E was not assessed due to the small sample size of homozygotes (n = 8). The presence of haplotype G was associated with increased weight (GLMM: $\beta_1$ = 0.712 ± 0.213, $p$ = 0.001; AM: $\beta_1$ = 0.688 ± 0.231, $p$ = 0.003), and there was a weak association in the same direction within dosage models (GLMM: $\beta_1$ = 0.712 ± 0.219, $\beta_2$ = 0.708 ± 0.785 $p$ = 0.004; AM: $\beta_1$ = 0.695 ± 0.234, $\beta_2$ = 0.540 ± 0.788, $p$ = 0.012).

In adults, August weight was associated with haplotypes C and G (Figure 5.4e-f, Appendix F). Weight was reduced in adults carrying haplotype C (GLMM: $\beta_1$ = –0.636 ± 0.203, $p$ = 0.0018; AM: $\beta_1$ = –0.618 ± 0.225 $p$ = 0.006) (Figure 5.4e), and this association was also marginally significant in the analysis of dosage (GLMM: $\beta_1$ = –0.627 ± 0.208, $\beta_2$ = –0.752 ± 0.673, $p$ = 0.0075; AM: $\beta_1$ = –0.632 ± 0.227, $\beta_2$ = –0.310 ± 0.670, $p$ = 0.021) (Figure 5.4f). Weight was greater in those individuals carrying haplotype G, though it was not significant within the AM (GLMM: $\beta_1$ = –0.586 ± 0.206, $p$ = 0.004; AM: $p$ = 0.11) (Figure 5.4e), and the association with dosage of haplotype G was marginal in the GLMM but not significant in the AM (GLMM: $\beta_1$ = –0.562 ± 0.211, $\beta_2$ = –0.952 ± 0.762, $p$ = 0.016; AM: $p$ = 0.27) (Figure 5.4f).

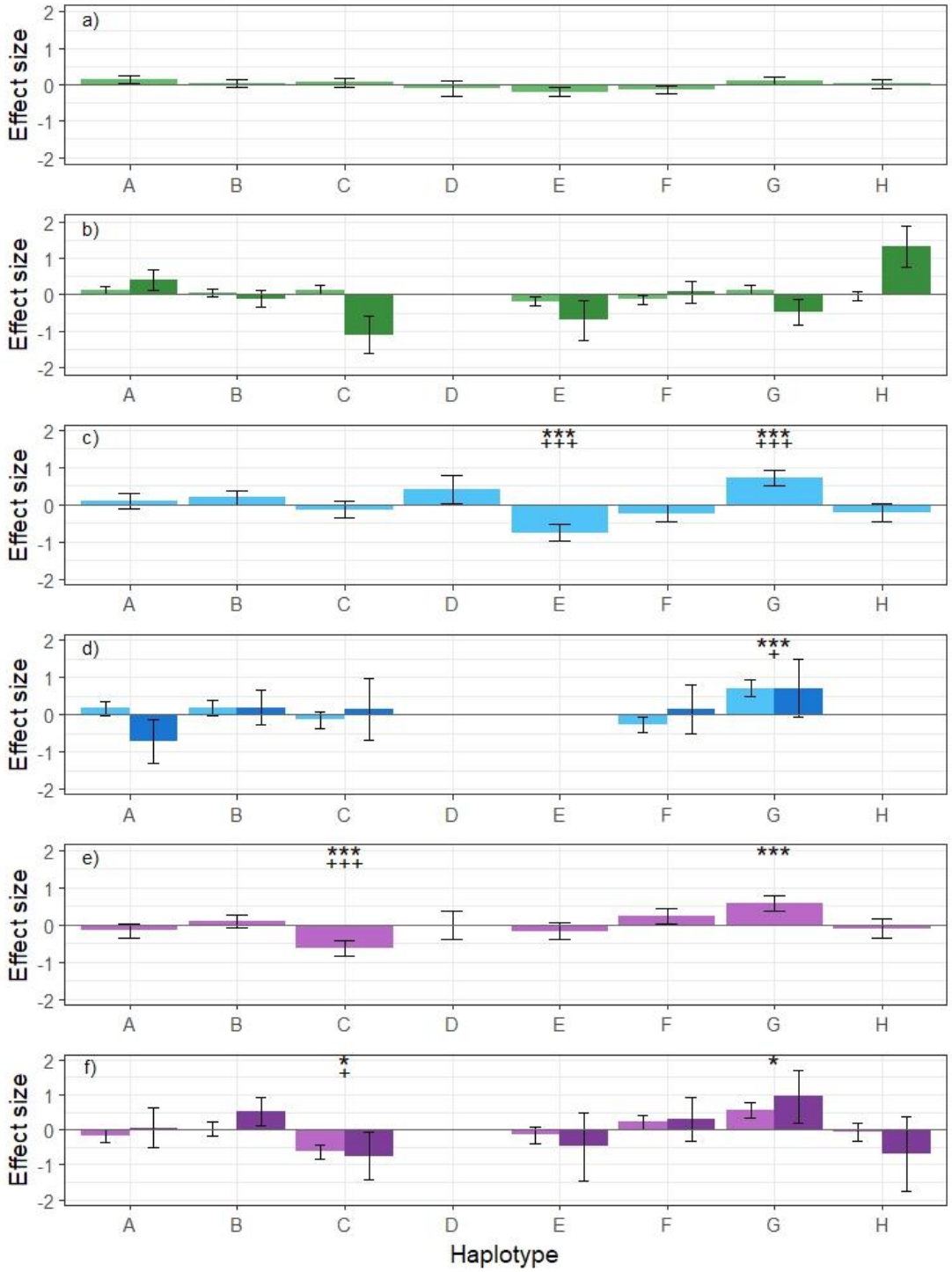



Figure 5.4. Effect sizes (and SE) from GLMMs of August weight in lambs for haplotype presence/absence (a) and dosage (b; light is one copy, dark is two copies); in yearlings for haplotype presence/absence (c) and dosage (d; light is one copy, dark is two copies); and in adults for haplotype presence/absence (e) and dosage (f; light is one copy, dark is two copies). The line at effect size 0 represents the intercept value for the haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold ($p < 0.0625$) is indicated as *** or +++, and marginal significance ($p < 0.05$) as * or +.

## 5.4.2 Immunoglobulins

### 5.4.2.1 Anti-*T. circ* IgA

There was no association between MHC class IIa heterozygosity and anti-*T. circ* IgA titres in lambs ($p = 0.612$) or adults ($p = 0.771$).

Anti-*T. circ* IgA titres were significantly higher in lambs with haplotype C ($\beta_1 = 0.092 \pm 0.031$, $p = 0.003$) and marginally lower in lambs with haplotype F ($\beta_1 = -0.066 \pm 0.027$, $p = 0.016$) (Figure 5.5a). When tested within an animal model, haplotype C was marginally significant after Bonferroni correction ($\beta_1 = 0.066 \pm 0.033$, $p = 0.043$) and haplotype F was not significant ($p = 0.27$). The association between anti-*T. circ* IgA and dosage was marginal for both haplotype C ($\beta_1 = 0.087 \pm 0.031$, $\beta_2 = 0.208 \pm 0.126$, $p = 0.007$) and haplotype F ($\beta_1 = -0.061 \pm 0.028$, $\beta_2 = -0.137 \pm 0.079$, $p = 0.03$) (Figure 5.5b), though neither haplotype C ($p = 0.108$) or haplotype F ($p = 0.5339$) were significant when tested in an AM (Appendix F). Within the GLMMs, there were no associations between haplotype E ($p = 0.054$) and haplotype G ($p = 0.356$), but when tested within an AM, haplotype E was marginally associated with increased IgA ($\beta_1 = 0.086 \pm 0.032$, $p = 0.0076$) and haplotype G with a decreased IgA ($\beta_1 = -0.074 \pm 0.003$, $p = 0.013$) (Figure 5.5a).

In adults, anti-*T. circ* IgA was lower in individuals carrying haplotype A ($\beta_1 = -0.105 \pm 0.033$, $p = 0.0017$) and haplotype F ($\beta_1 = -0.104 \pm 0.036$, $p = 0.0035$), higher with haplotype C ($\beta_1 = 0.127 \pm 0.037$, $p = 0.0006$), and there was a trend towards higher IgA with haplotype D ($\beta_1 = 0.164 \pm 0.067$, $p = 0.01$) (Figure 5.5c) when analysed within an GLMM. Within GLMMs, the associations remained within the dosage models for haplotype A ($\beta_1 = -0.092 \pm 0.034$, $\beta_2 = -0.232 \pm 0.092$, $p = 0.002$) and haplotype C ($\beta_1 = 0.128 \pm 0.038$, $\beta_2 = -0.114 \pm 0.129$, $p = 0.002$), but dosage of haplotype F was marginal ($\beta_1 = -0.096 \pm 0.037$, $\beta_2 = -0.202 \pm 0.113$, $p = 0.009$). None of these effects were significant when analysed within an AM (Figure 5.5d, Appendix F).

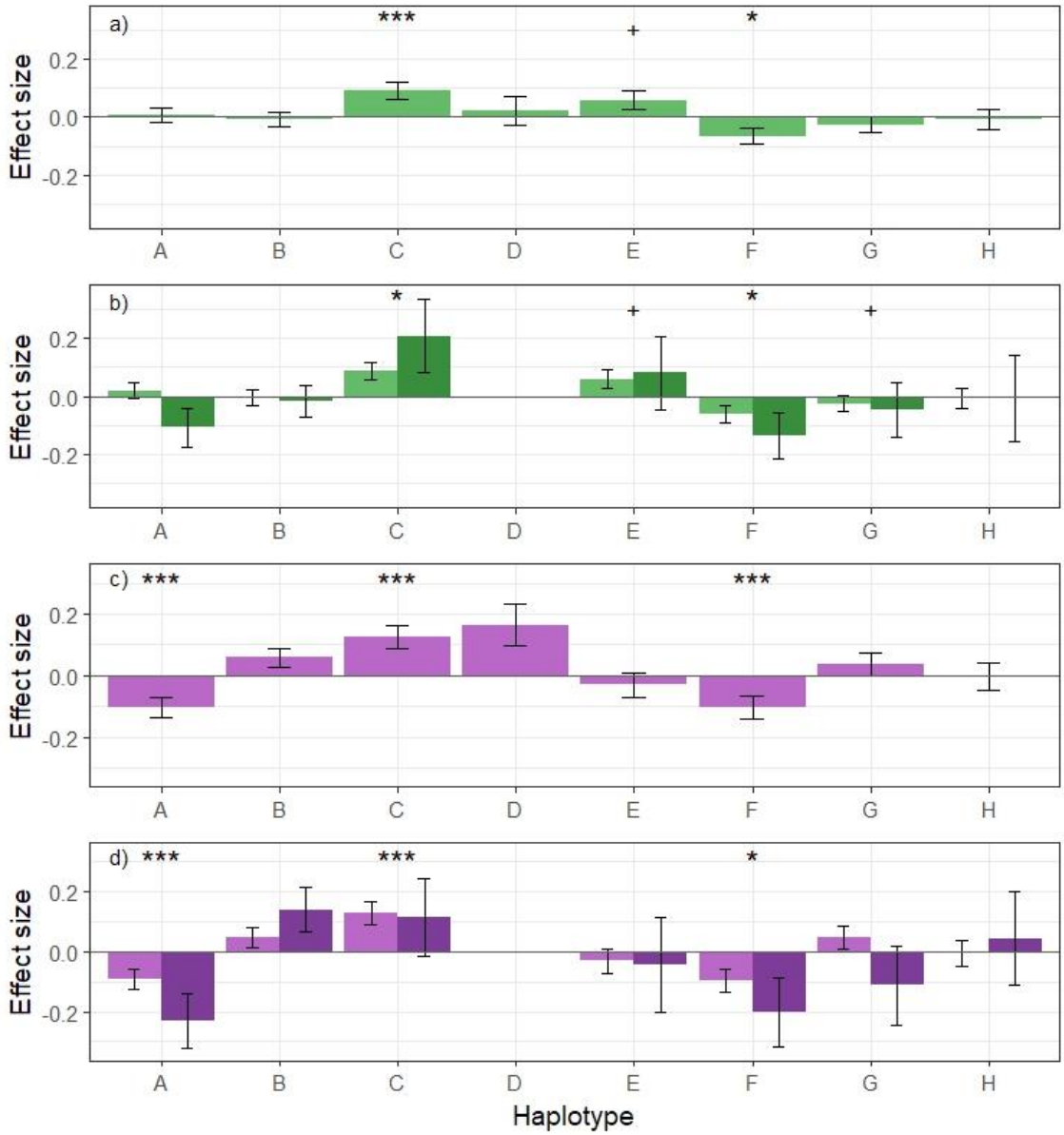



Figure 5.5. Effect sizes (and SE) from GLMM of associations between anti-*T. circ* IgA titres and a) presence of haplotypes in lambs, b) dosage in lambs where light is one copy and dark is two copies, c) presence of haplotype in adults and d) dosage in adults (light is one copy, dark is two copies). The line at effect size 0 represents the intercept value for then haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold ($p < 0.0625$) is indicated as *** or +++, and marginal significance ($p < 0.05$) as * or +. Note sample sizes were too small to analyse dosage of haplotype D in dosage analyses.

#### 5.4.2.2 Anti-*T. circ* IgE

There was no association between heterozygosity status and anti-*T. circ* IgE titres in lambs ($p = 0.798$) or adults ($p = 0.807$).

In lambs, anti-*T. circ* IgE titres tended to decline with haplotype A ($\beta_1 = -0.013 \pm 0.006$, $p = 0.04$) and increase with haplotypes D ($\beta_1 = 0.026 \pm 0.012$, $p = 0.032$) and F ($\beta_1 = 0.016 \pm 0.007$, $p = 0.0149$) within the GLMMs, although none of these models were significant after Bonferroni correction (Figure 5.6a, Appendix F). The haplotype F association was also marginal within the AM ($\beta_1 = 0.016 \pm 0.007$, $p = 0.024$), though associations were not present within the AM for haplotypes A ($p = 0.114$) or D ($p = 0.058$) (Figure 5.6a, Appendix F). Dosage of haplotype F tended toward increased anti-*T. circ* IgE titres, but only within the GLMM and not the AM (GLMM: $\beta_1 = 0.016 \pm 0.007$, $\beta_2 = 0.022 \pm 0.020$, $p = 0.049$; AM: $p = 0.079$) (Figure 5.6b, Appendix F).

Anti-*T. circ* IgE was lower in adults possessing haplotype A ($\beta_1 = -0.091 \pm 0.024$, $p = 0.0002$) and haplotype G ($\beta_1 = 0.081 \pm 0.027$, $p = 0.0024$), higher in adults possessing haplotype B ($\beta_1 = 0.105 \pm 0.023$, $p = 0.0002$) and tended to increase with haplotype D ($\beta_1 = -0.013 \pm 0.048$, $p = 0.027$) (Figure 5.6c). When analysed within an AM, the effect of haplotype B remained significant ($\beta_1 = 0.079 \pm 0.026$, $p = 0.002$), but those of haplotypes D ($\beta_1 = 0.123 \pm 0.056$, $p = 0.027$) and G ($\beta_1 = -0.069 \pm 0.029$, $p = 0.019$) were only marginal after Bonferroni correction and haplotype A was not significant ($p = 0.054$) (Fig 3c, Appendix F). Dosage of haplotype A ($\beta_1 = -0.088 \pm 0.025$, $\beta_2 = -0.119 \pm 0.067$, $p = 0.0007$), B ($\beta_1 = 0.077 \pm 0.024$, $\beta_2 = 0.149 \pm 0.053$, $p = 0.0004$) and G ($\beta_1 = -0.065 \pm 0.027$, $\beta_2 = -0.323 \pm 0.094$, $p = 0.0003$) were significant in the GLMM, but only haplotype B remained significant in an AM ($\beta_1 = 0.077 \pm 0.026$, $\beta_2 = 0.133 \pm 0.055$, $p = 0.005$), although haplotype G was marginal ($\beta_1 = -0.066 \pm 0.029$, $\beta_2 = -0.191 \pm 0.092$, $p = 0.023$) (Figure 5.6d). Dosage of haplotype D was not assessed due to the small sample size of the homozygotes (Table 5.3, $n_2 = 1$).

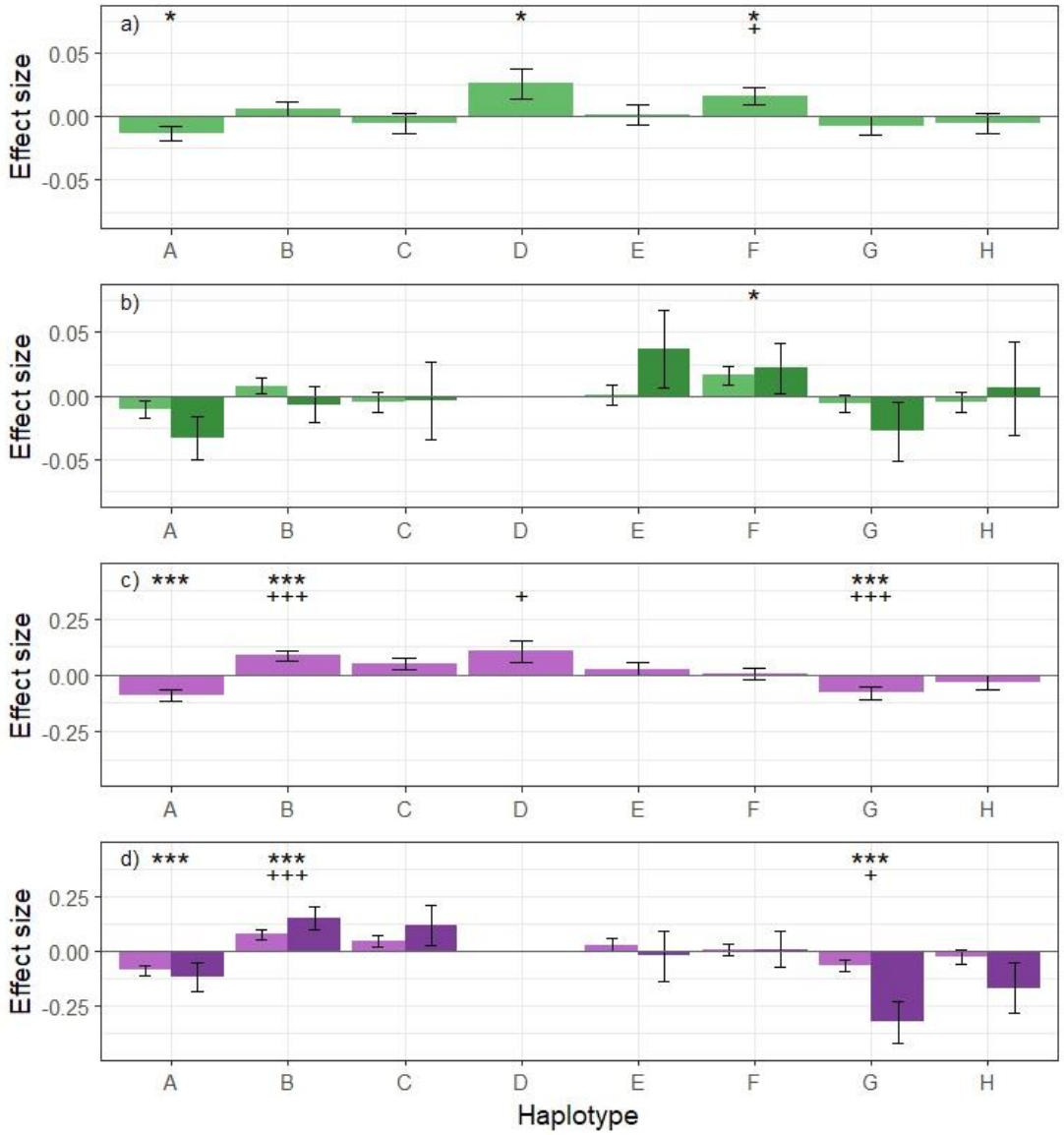



Figure 5.6. Effect sizes (and SE) from GLMM of associations between anti-*T. circ* IgE and a) presence of haplotypes in lambs, b) dosage in lambs where (light is one copy, dark is two copies), c) presence of haplotype in adults and d) dosage in adults (light is one copy, dark is two copies). The line at effect size 0 represents the intercept value for then haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold ($p < 0.0625$) is indicated as *** or +++, and marginal significance ($p < 0.05$) as * or +. Note sample sizes were too small for haplotype D in dosage analyses.

#### 5.4.2.3 Anti-*T. circ* IgG

There was no association between MHC class IIa heterozygosity and anti-*T. circ* IgG titres in lambs ($p = 0.676$) or adults ($p = 0.533$).

Lambs carrying haplotype E tended to have increased anti-*T. circ* IgG titres ($\beta_1 = 0.031 \pm 0.012$, $p = 0.0078$), which remained marginal within the AM ($\beta_1 = 0.029 \pm 0.012$, $p = 0.016$) (Figure 5.7a). Dosage of haplotype E was marginal within the GLMM, but significant within the AM (GLMM: $\beta_1 = 0.026 \pm 0.012$, $\beta_2 = 0.155 \pm 0.048$, $p = 0.0008$; AM: $\beta_1 = 0.025 \pm 0.012$, $\beta_2 = 0.157 \pm 0.047$, $p = 0.001$), with homozygous haplotype E having the highest anti-*T. circ* IgE titres (Figure 5.7b). Haplotype B tended to be associated with decreased anti-*T. circ* IgG titres ($\beta_1 = -0.021 \pm 0.009$, $p = 0.028$), as did haplotype G ($\beta_1 = -0.022 \pm 0.011$, $p = 0.041$), though only the haplotype G association held within the AM ($\beta_1 = -0.029 \pm 0.011$, $p = 0.009$) (Figure 5.7a & b). Neither dosage of haplotype A nor dosage of haplotype G were associated with anti-*T. circ* IgG titres (Appendix F).

In adults, anti-*T. circ* IgG was reduced with haplotype A, although the association was weak within the AM (GLMM: $\beta_1 = -0.033 \pm 0.011$, $p = 0.0045$, AM: $\beta_1 = -0.038 \pm 0.013$, $p = 0.046$) and also reduced with haplotype F (GLMM: $\beta_1 = -0.043 \pm 0.012$, $p = 0.0005$, AM: $\beta_1 = -0.038 \pm 0.013$, $p = 0.0041$) (Figure 5.7c). Haplotype E was marginally associated with increased anti-*T. circ* IgG (GLMM: $\beta_1 = 0.033 \pm 0.013$, $p = 0.013$, AM: $p = 0.011$) (Figure 5.7c). Dosage of haplotype F was significant within the GLMM, though marginal in the AM (GLMM: $\beta_1 = -0.041 \pm 0.013$, $\beta_2 = -0.060 \pm 0.039$, $p = 0.002$, AM: $\beta_1 = -0.038 \pm 0.013$, $\beta_2 = -0.043 \pm 0.040$, $p = 0.040$). Associations with dosage were only marginal within both the GLMM and AM for haplotype E (GLMM: $\beta_1 = 0.032 \pm 0.014$, $\beta_2 = 0.053 \pm 0.055$, $p = 0.044$, AM: $\beta_1 = 0.025 \pm 0.015$, $\beta_2 = -0.073 \pm 0.054$, $p = 0.030$), but only within the GLMM for haplotype A (GLMM: $\beta_1 = -0.032 \pm 0.012$, $\beta_2 = -0.039 \pm 0.032$, $p = 0.017$, AM: $p = 0.136$), and G (GLMM: $\beta_1 = 0.008 \pm 0.013$, $\beta_2 = 0.116 \pm 0.045$, $p = 0.034$, AM: $p = 0.271$) (Figure 5.7d, Appendix F).

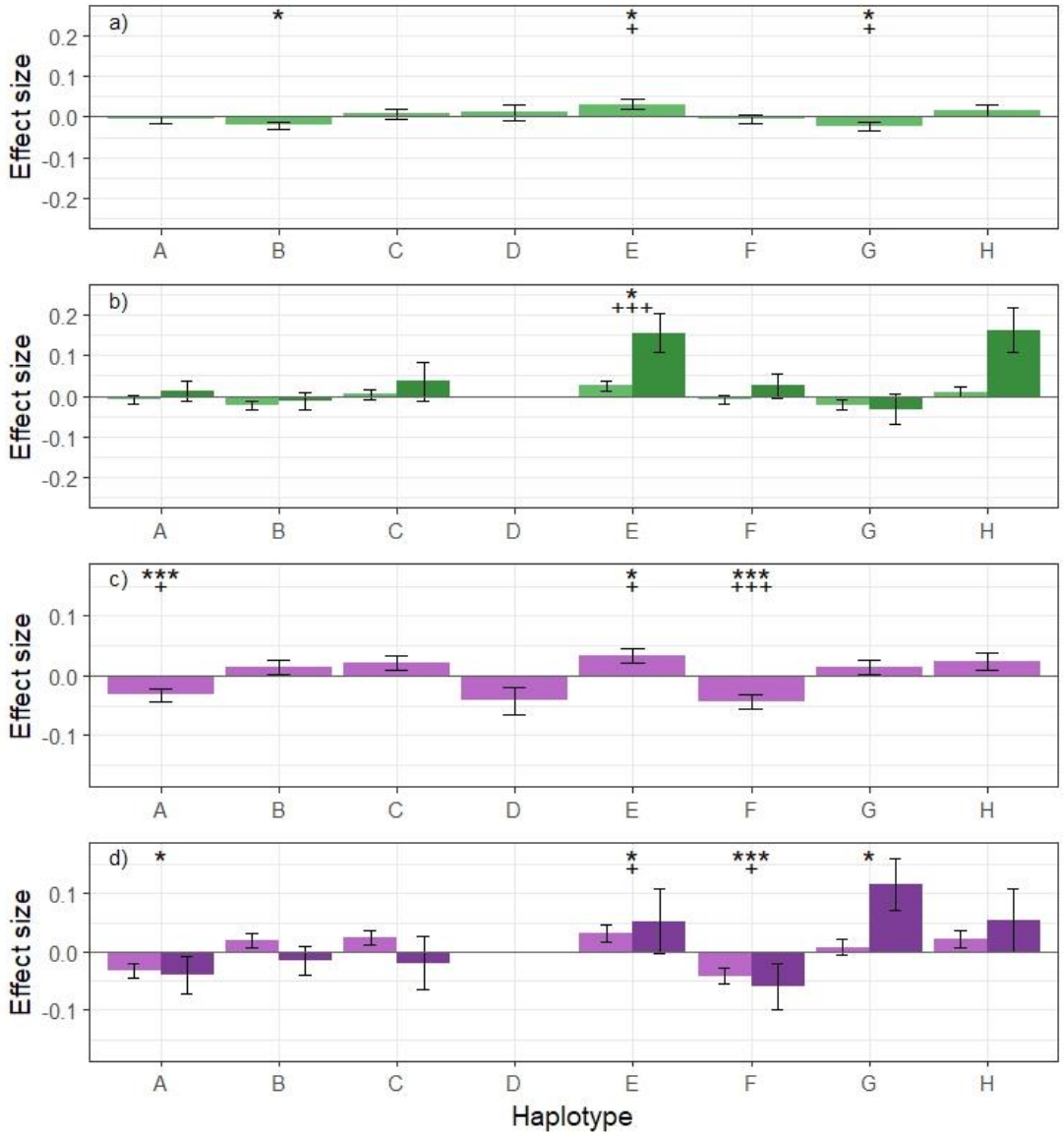



Figure 5.7. Effect sizes (and SE) from GLMM of associations between anti-*T. circ* IgG and a) presence of haplotypes in lambs, b) dosage in lambs where (light is one copy, dark is two copies), c) presence of haplotype in adults and d) dosage in adults (light is one copy, dark is two copies). The line at effect size 0 represents the intercept value for then haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold ($p < 0.0625$) is indicated as *** or +++, and marginal significance ($p < 0.05$) as * or +. Note sample sizes were too small for haplotype D in dosage analyses.

### 5.4.3 FEC

There was no association between MHC class IIa heterozygosity and FEC in lambs ($p$ = 0.529). Nor were there any associations between FEC and heterozygosity in yearling females ($p$ = 0.366), yearling males ($p$ = 0.572), adult females ($p$ = 0.804) or adult males ($p$ = 0.936).

In lambs, the presence of haplotype A was not significant in the GLMM but was marginal in the AM (GLMM: $\beta_1$ = 0.133 ± 0.063, $p$ = 0.073; AM: $\beta_1$ = 0.084 ± 0.051, $p$ = 0.023). Dosage of haplotype A was associated with increased FEC ($\beta_1$ = 0.044 ± 0.065, $\beta_2$ = 0.591 ± 0.161, $p$ = 0.0004), with haplotype A homozygotes having the highest FEC, though this effect did not hold within the AM ($p$ = 0.10) (Figure 5.8b). There was a weak positive association between the presence of haplotype B and lamb FEC ($\beta_1$ = 0.115 ± 0.059, $p$ = 0.046), though this did not hold within the AM ($p$ = 0.37). It should be remembered that FEC was modelled with a binomial error structure in GLMMs, but with a Gaussian error structure in AMs using log(FEC +50).

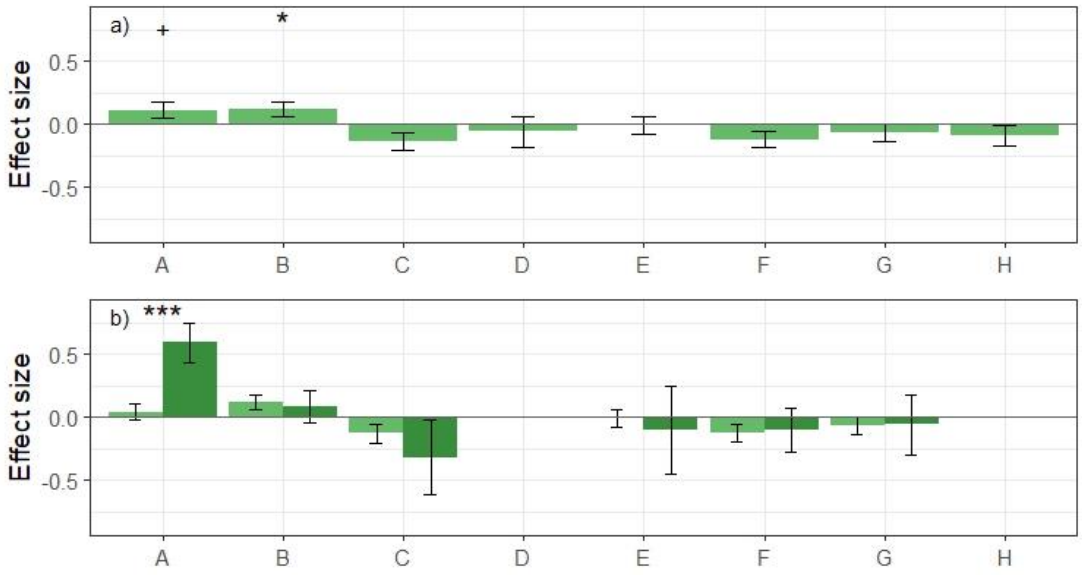



Figure 5.8. Effect sizes (and SE) from LMM of associations between mean Strongyle FEC in lambs as either a) presence/absence of the haplotype, or b) dosage of the haplotype (light is one copy, dark is two copies). The line at effect size 0 represents the intercept value for then haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in LMMs (asterisks) and AM (cross) at the Bonferroni threshold (α = 0.0625) is indicated as *** or +++, and marginal significance (< 0.05) as * or +. Note sample sizes were too small for haplotype D or H in dosage analyses.

Yearling FEC was analysed in separate models for males and females, and sample sizes were too small to analyse dosage effects. No associations between FEC and haplotypes were identified in yearling females (Figure 5.9a). Yearling males carrying haplotype A tended to have increased FEC in the GLMM ($\beta_1$ = 0.300 ± 0.149, $p$ = 0.042) and AM ($\beta_1$ = 0.297 ± 0.115, $p$ = 0.011) (Figure 5.9b).

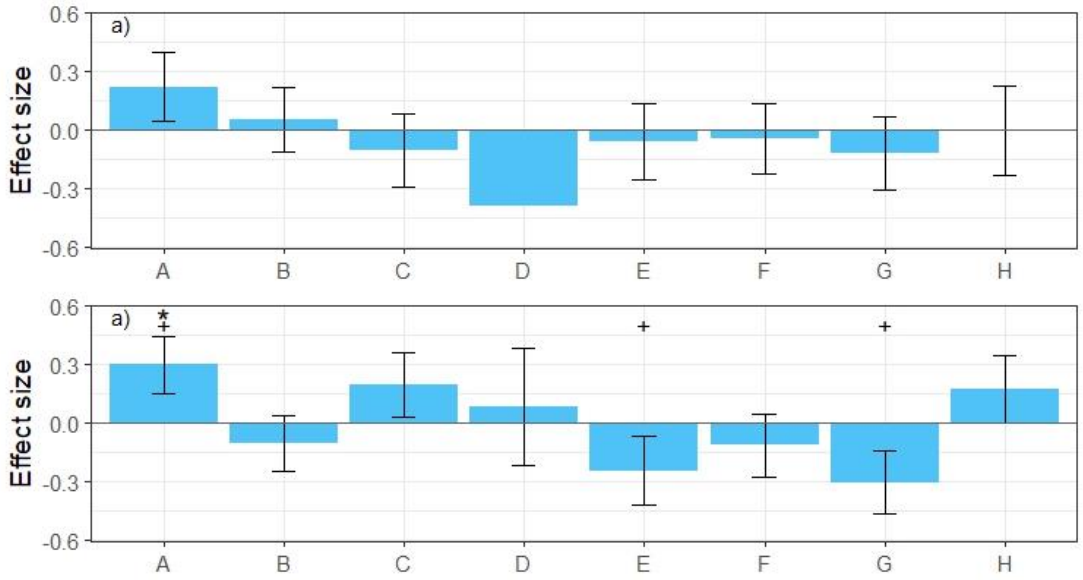



Figure 5.9. Effect sizes (and SE) from GLMM of associations between mean FEC and presence of haplotypes in a) yearling females and b) yearling males. Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold (α = 0.0625) is indicated as *** or +++, and marginal significance (< 0.05) as * or +. Note that sample sizes were too small to carry out dosage analyses.

In adult females, there were no associations between any haplotype and FEC when considering either presence or dosage effects within a GLMM (Figure 5.10a; Appendix F). When considered in an AM, there was a tendency for dosage of haplotype C to be associated with increased FEC ($\beta_1 = 0.028 \pm 0.059$, $\beta_2 = 0.513 \pm 0.203$, $p = 0.041$) in adult females, such that homozygotes for haplotype C had increased FEC, although the sample size is only $n_2 = 21$ (Figure 5.10c, Appendix F). In adult males, on the other hand, the presence of haplotype C was associated with decreased FEC in the GLMM ($\beta_1 = -0.363 \pm 0.156$, $p = 0.025$), which held in the AM ($\beta_1 = -0.355 \pm 0.119$, $p = 0.003$) (Figure 5.10b). Dosage was not assessed for males with haplotype C due to the small sample size of homozygotes (n = 8). Only haplotypes A, B and F had sufficient sample sizes analyse dosage, but no associations were identified (Appendix F).

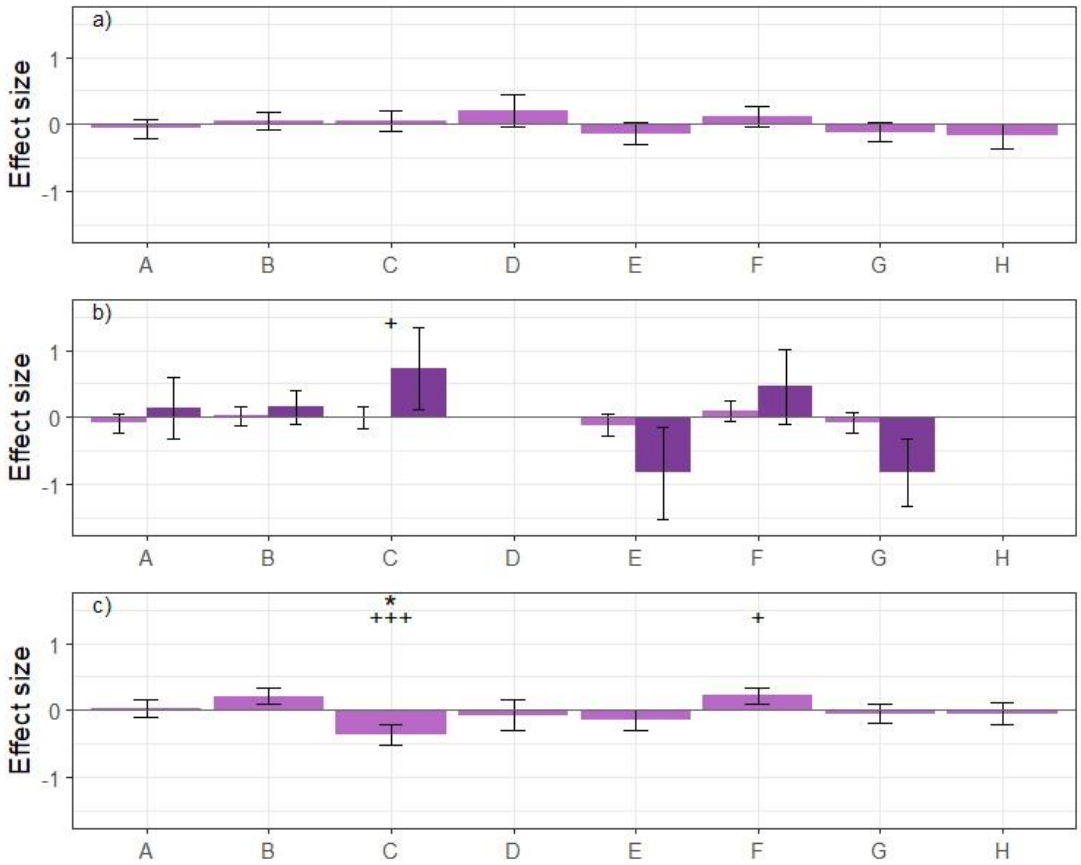



Figure 5.10. Effect sizes (and SE) from GLMM of associations between mean FEC and presence of haplotypes in a) adult females and b) adult males, and c) dosage of haplotypes in adult females. The line at effect size 0 represents the intercept value for then haplotype-absent category (i.e. 0 in presence/absence or dosage). Significance in GLMMs (asterisks) and AM (cross) at the Bonferroni threshold ($\alpha = 0.0625$) is indicated as *** or +++, and marginal significance (< 0.05) as * or +. Note that sample sizes were too small for haplotypes D and H in females and all haplotypes in males to carry out dosage analyses.

## 5.5 Discussion

To begin to understand the nature of the evolutionary processes acting to maintain MHC variation within this population, the relationships between the MHC class IIa haplotypes and phenotypes related to health were investigated within the Soay sheep population of St. Kilda. The haplotypes of individuals born within Village Bay between 1977 and 2012 were compared with August weight ($n_{lamb}$ = 1497, $n_{yearling}$ = 866, $n_{adult}$ = 928), anti-*T. circumcincta* IgA, IgE and IgG titres ($n_{lamb}$ = 1578, $n_{adult}$ = 1142), and FEC ($n_{lamb}$ = 1620, $n_{yearling}$ = 433, $n_{adult}$ = 620). There was no effect of MHC class IIa heterozygosity *per se* on any of the traits investigated, but there were associations between specific haplotypes and traits within sex-age classes, which are summarised in Figure 5.11. From a highly conservative perspective, where an association is considered significant at the Bonferroni threshold within an animal model, five haplotypes (B, C, E, F and G) were associated with variation in August weight, IgE titres, IgG titres and FEC in yearlings and adults. To my knowledge, this study represents the largest analysis of phenotypic associations of a free-living, non-human population, and the first to have sufficient sample sizes across different age classes to enable robust analyses within sex and age classes. Additionally, it is the first study to account for relatedness among individuals and show that relationships between haplotypes and phenotypes may be confounded by population stratification, generating false positives.

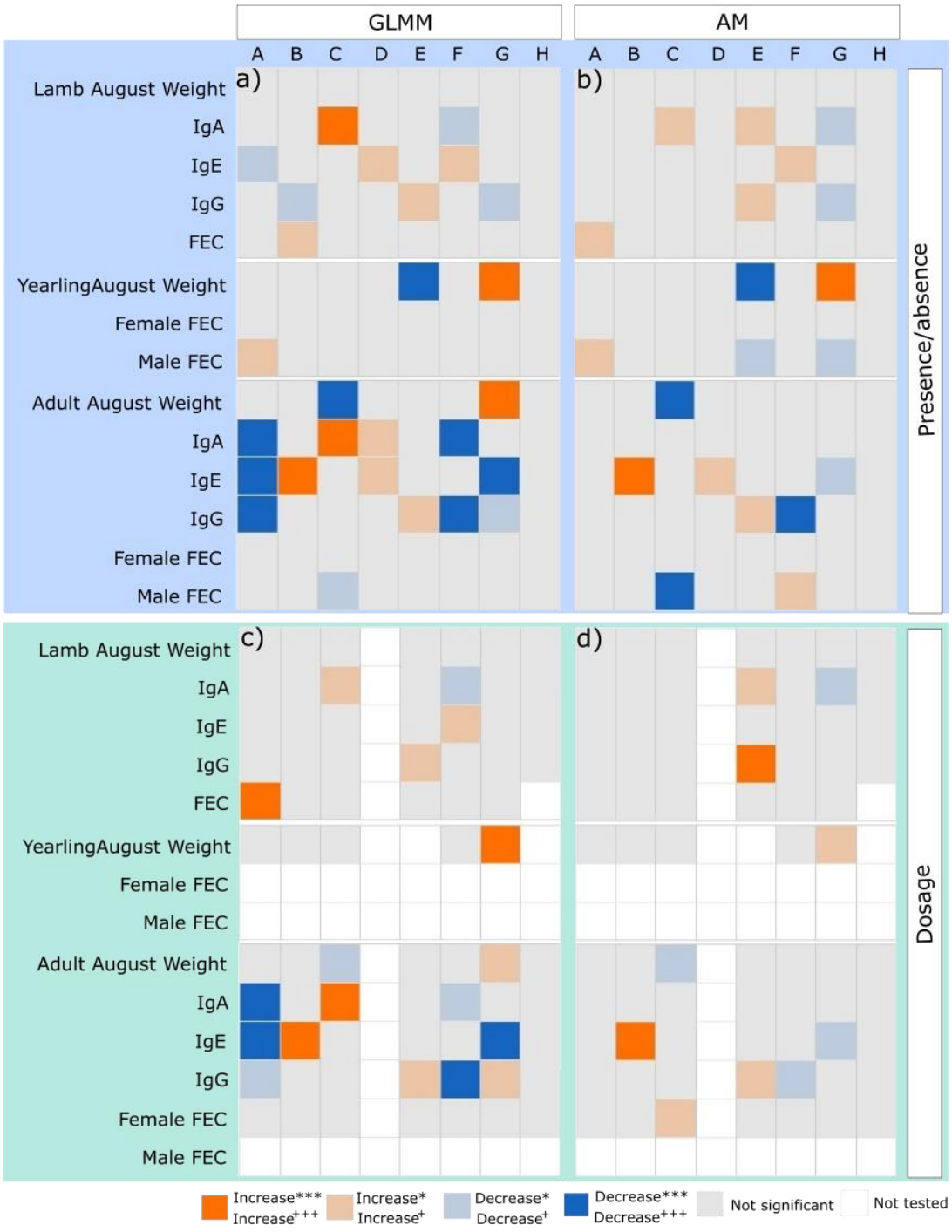



Figure 5.11. Summary plot showing associations for each age (and sex) category between haplotypes (columns) and traits (rows) for presence/absence-of-haplotype GLMMs (a), presence/absence-of-haplotype AMs (b), dosage GLMMs (c) and dosage AMs (d). Note that models which were not tested due too small sample sizes (homozygotes < 10) are shown in white.

### 5.5.1 Variation in associations across sex and age classes

Due to the large sample sizes in this study, associations between haplotypes and traits were tested within different sex and age classes. The costs associated with parasite infections, and thus the selection imposed by parasites on immunogenetic variation, may vary over ontogeny (Eizaguirre, Lenz & Eklund 2010). Parasites exert strong selection pressure on the Soay sheep (Gulland 1992; Coltman *et al.* 2001a; Wilson *et al.* 2004; Craig *et al.* 2006; Hayward *et al.* 2011), particularly on lambs (Hayward *et al.* 2011), and burdens and species composition vary with age (Craig *et al.* 2006; Hayward *et al.* 2009). Additionally, the acquired immune response develops over the first year of life (Stear, Strain & Bishop 1999b), and heritability of immunoglobulin measures increases with age (Strain *et al.* 2002; Davies *et al.* 2006, A. Sparks, unpub; Appendix F). The increase in plasma immunoglobulin measures between lambs and adults described here reflects this (Figure 5.2). If there is variation amongst haplotypes in their ability to present peptides from different species, their relationships with health and fitness may differ with age. The GLMM analyses carried out here show few robust associations (i.e. significant after Bonferroni correction; replicated in presence-absence and dosage tests) between haplotypes and traits in lambs (Figure 5.11), but more and stronger associations were found in older age classes. Within the more conservative sex-age specific animal models, despite high costs of parasitism and selection for parasite resistance in lambs (Hayward *et al.* 2011), only weak relationships were identified between class IIa haplotypes predicted to be involved in presenting peptides from gastrointestinal nematodes, and a measure of parasite burden and fecundity, FEC (Figure 5.11).

### 5.5.2 Relative lack of associations with FEC and comparison to earlier work

The MHC class IIa is involved in the presentation of peptides from gastrointestinal nematodes to the immune system (Murphy *et al.* 2012), and gastrointestinal nematodes are implicated in balancing selection maintaining MHC class IIa diversity (Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). In the present study, few associations were identified between MHC class IIa haplotypes and FEC, and only the presence /absence association between haplotype C and reduced FEC in adult males was significant within the AM after Bonferroni correction (Figure 5.11b). Apart from the stringent significance thresholds applied and the inclusion of these tests within an AM, several features of the FEC data may contribute to these results. The

measure of FEC used here is a very crude measure of parasite burden, both in terms of the way it is measured and what the measure actually represents in terms of species and burden. Because of the multiplier used in the method, the modified McMasters method (MAFF 1986) of estimating FEC as used here generates essentially categorical egg counts (i.e. 0, 100, 200, 300, etc), which combined with overdispersion makes modelling FEC challenging. Also, individuals with 0 eggs per gram may be true zeros and have no strongyle worms, or may have a very low burden. There may also be a very complicated relationship between FEC and the actual worm burden of an individual, for example in naturally infected Scottish Blackface sheep, Stear *et al.* (2006) found that high FEC was associated with a wider range of strongyle species. Therefore, increased FEC may be confounded with increased species diversity, and if there is variation amongst haplotypes in their ability to present peptides from different worm species, we may not expect to detect a direct association between haplotypes and FEC.

The lack of strong associations between FEC and MHC haplotypes is in contrast to an earlier study of MHC in the same population of Soay sheep (born between 1988 and 1994) which analysed genotypes of the DRB1-linked microsatellite OLADRB (Paterson *et al.* 1998). A positive association between FEC and OLADRB allele 257 bp was identified in lambs (predominantly haplotype B in this study; see Appendix C) and a negative association between FEC and OLADRB allele 263 bp in yearlings (haplotypes A and G here) and allele 267 bp (haplotype A here; see Appendix C) (Paterson *et al.* 1998). Whilst these differences may be due to different genotyping and analysis methodologies, it is likely that the previous result was a type I error. The loci under investigation differ between these studies. The OLADRB microsatellite alleles are shared across multiple class IIa haplotypes analysed here (Chapter 2), and there was probably a high genotyping error rate in the OLADRB microsatellite calling which has led to a number of OLADRB genotypes being associated with multiple diplotypes (see Appendix C). Additionally, the inclusion of maternal ID in both the GLMMs and AMs as a random effect here, but not within the previous study (Paterson *et al.* 1998), may have removed some variation in FEC of lambs that can be attributed to maternal effects, either genetic or environmental. The addition of 19 years more data in this study, compared to the previous Paterson et al. (1998) study likely eliminated spurious associations. There is a substantial difference in sample size and sample years between the study by Paterson *et al.* (1998) (data collected between 1985-1994) and the present study (data collected between 1985-2012). When the analyses carried out here

were re-run using only lambs born before 1995, and therefore a similar subset of individuals as analysed by Paterson *et al.* (1998), there was a tendency for individuals carrying haplotype B to have increased FEC (data not shown, p=0.045), which is consistent with the association found previously between OLADRB 257 (predominantly haplotype B; Appendix C) and increased FEC. A similar comparative subset of yearlings sampled before 1995 failed to detect associations between haplotypes and FEC, and was therefore unable to replicate the previous associations of Paterson *et al.* (1998) between OLADRB 263 and reduced FEC or OLADRB 267 and increased FEC. However, the OLADRB 263 allele is associated with both haplotypes A and G, and OLADRB 267 is also associated with haplotype A. Haplotypes A and G share the same DRB1 allele (DRB1*0101 Chapter 2), and Paterson (1998) showed that OLADRB 263 and 267 share the same DRB1 sequence. In carrying out these analyses using finer-scale resolution of the MHC haplotypes using many more years of data and a much more conservative statistical approach, spurious associations between MHC haplotypes and FEC have likely been removed.

The analyses carried out here did not incorporate any potential temporal variation in associations between haplotypes and traits. The evolutionary mechanisms of NFDS and FS both act at a temporal scale, resulting in allele (or haplotype) frequency changes caused by changes in selection pressure from pathogens. If either or both of these mechanisms were operating rapidly and within the timeframe of this study (27 years), associations between haplotypes and phenotypes would be expected to change, and by analysing the data in a single model, such variation would have been undetectable. It would therefore be interesting to look for such temporal variation in subsequent analyses, by, for example, looking at relationships between haplotypes and phenotypes (and fitness) over shorter time frames.

### 5.5.3 Haplotype dosage effects are consistent with presence/absence associations

Where it was possible to test for associations between traits and the dosage of a haplotype (all age, sex and haplotype classes containing at least 10 individuals) and an association was detected, the direction of the associations always reflected those identified within presence/absence models, and in most instances, homozygotes of the haplotype under investigation (dosage = 2) were similar or more extreme than heterozygotes of that haplotype (dosage = 1). For example, adult anti-*T. circ* IgE titre

was significantly increased in both GLMMs and AMs when an individual was heterozygous or homozygous for haplotype B, compared to individuals which did not carry haplotype B (Figure 5.6d); this mirrors the association identified in the presence/absence model (Figure 5.6c). Some associations that were detected in presence/absence models became non-significant when considering dosage, which was probably a consequence of loss of power due reduced sample sizes within heterozygous and homozygous categories compared with the combined category. There were four instances, however, where the associations were more pronounced in homozygotes of the haplotype under investigation, compared to heterozygotes or individuals without the haplotype. Adult anti-*T. circ* IgE titres were 17.4 times lower in haplotype G homozygotes than haplotype G heterozygotes (Figure 5.6d; LM***, $AM^{+}$). Lamb anti-*T. circ* IgG was 1.45 times higher in haplotype E homozygotes compared to haplotype E heterozygotes (Figure 5.7b; $GLMM^{ns}$, $AM^{+++}$). Lamb FEC was 1.07 times higher in haplotype A homozygotes compared to haplotype A heterozygotes (Figure 5.8b; GLMM***, $AM^{ns}$), and haplotype C homozygous adult females had a 1.14 times increase in FEC compared to haplotype C heterozygotes (Figure 5.10b; $GLMM^{ns}$, $AM^{+}$). Only the association between haplotype E and anti-*T. circ* IgG in lambs withstands Bonferroni correction and inclusion in an animal model, but these observations demonstrate that there are instances where homozygotes of a particular haplotype may deviate from the mean of a trait more extremely than heterozygotes of that haplotype.

### 5.5.4 Patterns of trait associations across haplotypes

If haplotypes are tightly associated with parasite resistance, we might expect to find patterns that link associations across traits within a haplotype. Within Soay sheep, at the phenotypic level, August weight is negatively correlated with FEC (Coltman *et al.* 2001a; Craig *et al.* 2008; Hayward *et al.* 2014b) and positively correlated with fitness (Milner *et al.* 1999; Jones *et al.* 2005), anti-*T. circ* IgA is negatively associated with FEC (Coltman *et al.* 2001b), anti-*T. circ* IgE is positively associated with FEC (Nussey *et al.* 2014), and anti-*T. circ* IgG is positively associated with survival (Nussey *et al.* 2014; Watson 2016); thus, a haplotype associated with increased August weight, increased IgA, decreased IgE, increased IgG and decreased FEC would be expected to relate to increased fitness. In this study, no haplotypes were associated with more than two traits within animal models after Bonferroni correction (Figure 5.11); however, a number of observations can be made by also considering the pre-Bonferroni correction trends across traits for a haplotype.

The patterns of associations tended to be opposite for haplotypes E and G (Figure 5.11), which may suggest opposite effects on an individual's fitness. Haplotype E was associated with decreased weight (yearlings), increased immunoglobulin measures (lamb IgA; lamb and adult IgE), and decreased FEC (yearling males). Haplotype G, on the other hand, was associated with increased weight (yearlings), decreased immunoglobulin measures (lamb IgA -dosage only, adult IgE, and lamb IgG) and decreased FEC (yearling males). In both these cases, the association with FEC was only observed within the animal model above the Bonferroni threshold.

A negative relationship between body weight and FEC has been observed within Soay sheep (Coltman *et al.* 2001a; Craig *et al.* 2008; Hayward *et al.* 2014b; null models in this study), and the potential trade-offs between mounting an immune response and growth are widely recognised (Lochmiller & Deerenberg 2000; Schmid-Hempel 2003), yet this study found that haplotypes C and possibly E were associated with both reduced weight and reduced FEC, although the associations with FEC should be regarded with caution due to the use of Gaussian error structure with log transformed FEC within the AM, rather than using a negative binomial error structure. Haplotype C was associated with high lamb and adult IgA titres in the GLMM but not the AM, which suggests that individuals with haplotype C tend to have high IgA titres although that variation is attributed to shared phenotypic variation amongst related individuals. Haplotype E was also associated with increased lamb IgA and adult IgG titres. This suggests that there may be a complex relationship between MHC class IIa haplotype, body weight, immunoglobulin measures and FEC, where a haplotype associated with a strong immune response can reduce growth and weight gain but does improve an individual's ability to control parasites, as measured by FEC. Nussey *et al.* (2014) found that total IgG was negatively and quadratically related to body weight in Soay sheep, and Watson (2016) found that anti-*T. circ* IgE and anti-*T. circ* IgG were positively but quadratically associated with growth in the 2011 cohort of Soay lambs. This may indicate that very high immunoglobulin measures, whilst effective at controlling FEC, may have a negative trade-off with other important life-history traits, such as weight-gain and growth. Given that the immunoglobulin measures are generally consistent in adults across an individual's lifetime (repeatability of IgA = 0.76, IgE = 0.74 and IgG = 0.53; A. Sparks, unpublished data) and that the immunoglobulin measures considered here seem to be cross-reactive to a wide range of nematode species and are not specific to *T. circumcincta* (D. Nussey & K. Watt, pers. comm.), some combination of genetics

and early life gastrointestinal nematode experience probably sets an individual's lifetime immunoglobulin titre trajectory. MHC diplotypes are likely to be involved in determining that trajectory as the acquired immune response is developed throughout early life.

### 5.5.5 The impact of applying animal models

Throughout this study, associations between haplotypes and traits were assessed using both standard linear mixed effects models (GLMM), from a family of statistical models widely use in studies of MHC associations in wild populations, and animal models (AM). The animal model framework includes phenotypic information from individuals of varying relatedness to estimate the additive genetic component of the trait by including the breeding value as a random effect within a mixed effect model (Wilson *et al.* 2010). By including this random effect, it is possible to account for relatedness between individuals and thereby account for variation in the trait that is due to additive genetic effects located throughout the genome. When analysed within an animal model, a number of associations between haplotypes and traits that were identified in standard GLMMs became only marginally significant (after Bonferroni correction) or were not detected at all (Figure 5.11). Most notably, a number of associations with IgA in adults were not present at all within the animal models and only marginally significant within lambs (Figure 5.11 compare panel (a) with panel (b)). The additive genetic variation for IgA was high in this study in lambs ($h^2$ = 40.0 % ± 5.6 %) and adults ($h^2$ = 60.5 % ± 4.6 %) (see Appendix F), and suggests that the variation that was being attributed to Class IIa haplotype variation within the GLMMs was actually due to genome-wide additive genetic variation. Indeed, a recent quantitative GWAS of the immunoglobulin data set used here found a large-effect QTL for IgA on chromosome 24 (A. Sparks, unpublished data).

This is the first time, to my knowledge, that associations between MHC variation and phenotypes have been analysed within an AM framework. By carrying out analyses using both standard GLMMs and AMs, it is clear that a number of apparently strong associations were lost after accounting for relatedness. When there are related individuals within a dataset, there may be an increased risk of detecting spurious results by not accounting for the fact. This risk may be especially high when the genome region under investigation is highly polymorphic, as with the MHC, since it may detect relatives with some efficiency.

There were four marginally significant associations between haplotypes and FEC that were identified only within the AM. These differences were potentially caused by the differences in model error structures – the GLMM used a negative binomial error structure and raw FEC measures, whereas the AM used a less-than-ideal Gaussian error structure on transformed log(FEC+50).

In experimental studies, it is possible to use full-sibling and half-sibling families to control for genome similarity, and incorporate family ID within mixed effect models or survival analyses to control for this problem (e.g. Kjøglum *et al.* 2008; Eizaguirre *et al.* 2012a; b). Studies conducted by sampling individuals from wild populations may sample relatives, but to date do not generally account for relatedness between individuals in their analyses. Analysis using an animal model, however, requires accurate estimates of relatedness, either through intensive sampling of individuals and genetic parentage analysis to generate a pedigree or from genome-wide SNP data (Berenos et al 2014). Where there is substantial additive genetic variation for a trait, by not considering relatedness, there may be an increased risk of generating false positive associations with MHC genotypes or haplotypes.

### 5.5.6 Evolutionary mechanisms

The simplest evolutionary model to explain MHC diversity is heterozygote advantage or overdominance (Takahata & Nei 1990; Hedrick 2002; De Boer *et al.* 2004), but this study found little support for this hypothesis in Soay sheep. Heterozygote advantage can be detected as operating when the mean fitness (or value of a trait likely to contribute to fitness) of the heterozygote is greater than that of either homozygote. Here, this possibility was tested in two ways. First, I compared whether being a heterozygote for any pair of MHC class IIa haplotypes conferred different trait values relative to being a homozygote for any haplotype. Secondly, for each haplotype three diplotype classes (containing 0, 1 and 2 copies of the focal haplotypes) were fitted and whether they were associated with different trait values (dosage analyses) was determined. In the latter analyses, heterozygote advantage would manifest as a higher trait values for the '1' class than for the '0' and '2' classes. Neither approach found any evidence that heterozygotes have higher values than homozygotes for any of the traits considered here, either in the GLMMs or in the AMs. Previous studies have detected heterozygote advantage in systems with a small number of MHC alleles or haplotypes (e.g. Chinook salmon – Arkush *et al.* 2002; Arctic charr – Kekäläinen *et al.* 2009; water

voles – Oliver *et al.* 2009b; mountain bushtail possum – Banks *et al.* 2010; New Zealand sea lion – Osborne *et al.* 2015) or in experimental systems where the number of alleles or haplotypes is artificially limited (Penn, Damjanovich & Potts 2002; McClelland *et al.* 2003). In these studies it is frequently noted that heterozygote advantage is only detected in multi-pathogen infection, presumably because different alleles offer resistance or susceptibility to different pathogens (McClelland *et al.* 2003; Stoffels & Spencer 2008).

In systems in which alleles cannot easily be attributed to specific loci, typical of studies in birds and fish (Appendix A), the number of alleles is often used as an indirect measure of heterozygosity. This assumes that individuals with larger numbers of alleles are more likely to be heterozygous at more loci, or, more generally, that the same principle of heterozygote advantage applies - that individuals with more alleles are able to present peptides from a broader range of pathogens. In this instance, studies find that increased allele number is associated with improved performance in a trait of interest (Brouwer *et al.* 2010; Dunn *et al.* 2013; Pavey *et al.* 2013) or that an intermediate number of alleles is optimal (Madsen & Ujvari 2006; Wegner 2008; McCairns *et al.* 2011; Hawley & Fleischer 2012), presumably because high individual allelic diversity could lead to self-reactivity (an autoimmune response) (Nowak *et al.* 1992). There are eight class IIa haplotypes in the Soay sheep (Chapter 2), which were analysed here. However, due to the fact that some alleles are common to multiple haplotypes and haplotype G was found to have multiple DQB2 alleles (Chapter 2), the number of alleles an individual carries varies between 5-6 in homozygotes and 8-11 in heterozygotes. However, there was no effect of allele number on any of the traits analysed within this chapter (data shown in Appendix F).

Considering whether a locus or haplotype is heterozygous or counting alleles may not be the best way of determining whether a more heterozygous individual is able to present a wider variety of pathogens. MHC alleles are thought to evolve through gene conversion (Lawlor *et al.* 1990; Ohta 1999) and for this reason can be very similar to one another in their nucleotide sequence and presumably their function. Additionally, alleles vary in their specificity, and there may be degeneracy in their ability to bind foreign peptides (Stoffels & Spencer 2008). A heterozygote which carries alleles that are functionally similar may not enjoy the selection advantage afforded by carrying functionally divergent alleles. The functionality of MHC alleles is not, however, easy to determine. If inter-haplotype MHC molecules can form, this could enhance the

functional capabilities of heterozygotes, although Ballingall *et al.* (in press) have shown that not all combinations of A and B alleles generate functional molecules. One way of addressing this question is to classify alleles according to the properties of the amino acids within the peptide binding region (Doytchinova & Flower 2005; Naugler & Liwski 2008) and perform a multivariate analysis (e.g. principal component analysis or co-inertia) to cluster similar alleles, or perhaps generate a dissimilarity score. This approach has often been successful in species with very large numbers of alleles (Schwensow *et al.* 2007; Huchard *et al.* 2010; Sepil *et al.* 2013a; Sepil, Lachish & Sheldon 2013b; Jones, Cheviron & Carling 2014; Lillie *et al.* 2015), in which statistical power may otherwise have been unachievable for an allele- or haplotype-based analysis like the present one. The ability to cluster haplotypes in the Soay sheep is currently limited by both the lack of knowledge of which sites are involved in peptide binding (see Chapter 3), and lack of sequence data for the *DQB* loci in haplotype D.

### 5.5.7 Large sample sizes have not entirely relieved the statistical challenges

The highly diverse nature of the MHC, whilst being what makes it interesting, makes it challenging to attain sufficient sample sizes to carry out statistical analyses (Hill 1998). By using a rapid and high-throughput SNP-based genotyping method (Chapter 3), this study was able to create one of the largest sample sizes available to any MHC study, with the exception of human studies, and yet, a number of dosage analyses, particularly for the rarest haplotypes D and H, could not be carried out due to small sample sizes. This highlights one of the key challenges of investigating the processes underlying MHC variation. A key theory underpinning the maintenance of diversity in this region is NFDS, in which rare alleles have a fitness advantage, causing them to increase in frequency within the population, when the fitness advantage of that allele decreases and reverts to another, now rare, allele (Bodmer 1972; Takahata & Nei 1990). The lack of associations between haplotypes D and H with the traits assessed here could be due to their rarity with the population compared to the alternative haplotypes, and thus a lack of statistical power. Within 5349 genotyped individuals representing the vast majority of individuals that were born or died within Village Bay between 1985 and 2012, only six homozygous haplotype D individuals were identified, although 308 were heterozygous for the haplotype (Chapter 4). When this sample was further pruned within statistical analyses to include only individuals with a known birth year and phenotypic data, as well as subdivided into age and sex classes (as for FEC analyses),

numbers of homozygotes became vanishingly small; thus, even with large numbers of individuals and 27 years of phenotypic data, testing negative frequency-dependent selection had low power for some haplotypes in these analyses. Additionally, if different pathogens have different associations with the same haplotypes or if alleles have weak effects within contemporary populations, the problem of large numbers of alleles and low sample sizes may make it even more difficult to detect associations (Apanius *et al.* 1997; Piertney & Oliver 2006).

Finally, I investigated the associations between haplotypes and traits, rather than between specific loci and traits. This approach was taken both for statistical tractability and because high LD among the Class IIa expressed loci suggests selection will indeed be on the combination of alleles contained in each haplotype. However, this approach does mean that the positive and negative effects of different alleles at different loci on a trait are essentially averaged across the haplotype. An alternative and more complex approach to analysis would be to consider the net effect of each allele at each locus separately (including potential interactions between loci), but this is beyond the scope of this thesis due to time constraints.

# 6. General Discussion

## 6.1 Summary

Studies of MHC evolution in wild systems rarely generate the locus-specific genotype data across multiple loci that are necessary to investigate which evolutionary mechanisms operate to maintain diversity (Spurgin & Richardson 2010). The primary aim of this thesis was to characterise the functional MHC class IIa diversity in a wild mammal and generate diplotype data for a large sample of individuals. A secondary aim was to investigate the evolutionary mechanisms underpinning the class IIa diversity by asking first, is balancing selection operating, second, is heterozygote advantage operating, and third, are there associations between haplotypes and phenotypes related to health? In chapter 2, I characterised the class IIa diversity and identified eight haplotypes. In chapters 3 and 4, I developed a method for haplotyping large numbers of individuals using KASP genotyping and generated haplotype data for what constitutes the largest MHC typing effort for any single population of a non-model species. In chapter 4, I show that balancing selection is likely operating to maintain more even allele frequencies than predicted under neutrality, and in chapter 5, show that heterozygote advantage is an unlikely mechanism for maintaining diversity within this population, but that there are a range of associations between haplotypes and weight, immunoglobulin titers and FEC. Whilst in this thesis I have only touched on the questions of how diversity is maintained in the Soay sheep, I have developed a large, high-quality data set of MHC haplotypes within a well-characterised population that should be valuable in exploring such questions in the future.

## 6.2 Class IIa haplotypes

By fully characterising the class IIa haplotypes of Soay sheep, I have shown that the variation across the class IIa region is actually limited to only eight haplotypes, but also that no single locus had unique alleles on all eight haplotypes. Many studies genotype a single class II locus (see Appendix A) with the expectation that, due to LD between the class II loci, each allele represents a single haplotype. In chapter 2, it is clear that no single locus is representative of the class IIa variation in the Soay sheep, i.e. no haplotype could be uniquely identified by genotyping a single locus. This could affect downstream association studies by reducing the ability to detect associations if an allele occurs on multiple haplotypic backgrounds which have different effects on the

trait of interest. Haplotypes A and G, as well as B and E have more than one allele in common (Chapter 2), and while no two trait associations were in opposite directions for these haplotypes, associations in the same trait was rarely identified for both haplotypes in each pair (Chapter 5). Multi-locus haplotyping of the MHC class II region is therefore valuable to downstream analyses.

The diplotype data set generated here represents the largest haplotype-based data set for the MHC in a non-model species. The relatively low diversity at the class IIa region in Soay sheep greatly facilitated haplotype characterisation, likely due to it being an island population with a lack of immigration, as well as the founder event in 1931 which established the studied population on the island of Hirta from only 85 fertile individuals translocated from Soay (Boyd 1953; Boyd *et al.* 1964). Only a handful of studies in non-model species have generated locus-specific genotypes for more than one locus and they are almost entirely restricted to mammals (e.g. Kennedy *et al.* 2011; Smith *et al.* 2011; Miyasaka *et al.* 2012; Gillet *et al.* 2013; Liu *et al.* 2013; Arbanasić *et al.* 2014; Ali *et al.* 2016; Ruan *et al.* 2016; Vasoya *et al.* 2016). Only Osborne *et al.* (2015) and Niskanen *et al.* (2014) carried out MHC association studies using haplotype information.

## 6.3 A SNP haplotyping method

Genotyping the MHC is notoriously challenging (Bernatchez & Landry 2003; Piertney & Oliver 2006; Babik 2010; Spurgin & Richardson 2010), and locus-specific genotyping methods rarely scale well for high-throughput genotyping. Indirect methods, such as denaturing gradient gel electrophoresis (DGGE), single-strand conformation polymorphism (SSCP) and reference strand-mediated conformation analysis (RSCA) generate allele-specific conformations following PCR which enables allele separation (Lenz *et al.* 2009a; Babik 2010). Such methods require relatively high initial optimisation, do not generate allele sequence information without subsequent sequencing, and are not particularly high-throughput (Babik 2010). Sequence-based genotyping (SBG), as used in chapter 2, is excellent for generating high-quality sequence data from locus-specific amplifications, but can require costly cloning of multi-allelic sequences which might be obtained from either heterozygous loci or multi-locus amplifications and therefore does scale up to high-throughput analyses (Babik 2010; Lighten *et al.* 2014). Next-generation sequencing (NGS) methods scale up to high-

throughput analyses with little optimisation and are particularly valuable for phasing multi-locus amplifications as each allele is sequenced in-phase (Babik 2010). However, PCR amplification of multiple very similar alleles can lead to the generation of low frequency artefact alleles which are detected by NGS, such as chimeras and base-mismatch errors (Lighten *et al.* 2014), which are not easily separated from true rare alleles (Babik 2010; Sepil *et al.* 2012; Lighten *et al.* 2014; Vasoya *et al.* 2016). Finally, an Illumina Immunochip containing approximately 200K polymorphisms (SNPs and insertion/deletions) has been developed for typing the human HLA and other immune genes (Cortes & Brown 2011), but extensive population data and a well assembled genome is required to develop such a panel, making it unrealistic for non-model species. There are advantages and pitfalls to all genotyping methods currently applied to MHC typing in non-model organisms (Babik 2010), and the choice of method will often be determined by the MHC variability and single-locus amplification availability for the species in question, as well as the number of samples and loci that require genotyping.

In characterising the class IIa region in Soay sheep, I made use of existing SBG methods for domestic sheep (Ballingall & Tassi 2010; Ballingall *et al.* 2015, 2017). SBG methods developed for domestic sheep were successful in characterising most of the variation in the MHC class IIa genes in Soay sheep (chapter 2). The existing IPD-MHC database of allelic diversity at the *DRB1* locus was valuable in that it enabled me to identify all *DRB1* alleles within the study population by amplifying only exon 2 from gDNA. A similar database of alleles at the *DQA* and *DQB* loci is not yet available in the IPD-MHC, and a number of exon 2 sequences did not match previously published full-length transcripts (Ballingall *et al.* 2015, 2017). Thus, costly and time-consuming cloning from cDNA was necessary to fully characterise allelic variation at these loci. This was not possible for the *DQB* loci in the rare haplotype, D, due to the multi-locus amplification of *DQB* loci, lack of homozygous sheep and problems amplifying full length transcripts. Nevertheless, novel alleles were identified at both *DQB1* and *DQB2* loci in haplotype D from heterozygous sequences amplified from gDNA, although the alleles could not be separated using these sequences. All haplotypes in the Soay sheep were defined, even though not all alleles were characterised in full.

SBG of class IIa loci was time-consuming and costly, and could not be scaled to the numbers of sheep desired for subsequent analyses, so I made use of existing SNP data for the Soay sheep (Bérénos *et al.* 2014; Johnston *et al.* 2016) to develop a rapid SNP

haplotyping method. Soay sheep had previously been genotyped on two SNP chips, 5805 individuals on the Ovine SNP50 BeadChip (Bérénos *et al.* 2014) and 188 key individuals on the Ovine Infinium HD SNP BeadChip (Johnston *et al.* 2016). SNPs selected for inclusion on SNP chips are typically evenly spaced throughout the genome, have high minor allele frequencies (MAF), and can be successfully genotyped in different populations (Kijas *et al.* 2012). The MHC region, on the other hand, is typified by tightly clustered and highly variable genes, with low MAF, and variations in the number and type of genes between haplotypes can cause SNP failures on haplotypes lacking the locus. Additionally, designing primers for such highly variable regions is very challenging. Therefore, the almost complete lack of SNPs in the class IIa region on the less dense Ovine SNP50 BeadChip and low density on the Ovine Infinium HD SNP BeadChip was unsurprising. I did find, however, that SNPs on the Ovine Infinium HD SNP BeadChip were capable of imputing the class IIa haplotypes, and so a panel of SNPs was selected for use in high-throughput genotyping in the Soay sheep. A panel of SNPs has the advantage of being relatively easy to genotype, suitable for high-throughput genotyping and does not suffer from artefacts, but it is susceptible to incorrect haplotype phasing due to genotyping errors, particularly allelic dropout which may be most prevalent in poor quality or quantity DNA (Gill 2001; Butler 2005). Stringent quality control measures, as applied in Chapter 4, should minimise any effects of genotyping and phasing errors, and the Soay sheep pedigree was particularly valuable here in enabling the inclusion of a Mendelian inheritance check as a quality control measure. As a method for genotyping MHC in a non-model organism, however, this panel of SNPs required high initial optimisation effort in terms of preliminary class IIa sequencing, and is not easily transferable to other breeds of sheep or other species without extensive preliminary sequencing of haplotypes. Development of such a method using existing SNPs for the MHC class I region may also not be feasible as the class I genes are even less well characterised and mapped in the sheep genome than class II genes.

## 6.4 Evidence for balancing selection

The Ewens-Watterson tests carried out in Chapter 4 provide good evidence that balancing selection is acting on the class IIa region in the Soay sheep as, irrespective of life history stage or the year of standing population tested, there was evidence of more

even allele frequencies than would be expected under neutrality, indicative of balancing selection. This test assumes that a population is at equilibrium and at a constant size (Garrigan & Hedrick 2003), but it is unlikely that these results were caused by demographic effects as the test was consistently significant across years within the standing population, and there is no evidence of such demographic effects generating even allele frequencies at neutral microsatellite loci (Paterson 1998; Charbonnel & Pemberton 2005).

That association between haplotypes and phenotypes were identified in Chapter 5 suggests that variation in haplotype function is reflected in phenotypic variation in traits related to individual health, which lends weight to the possibility that pathogen-mediated selection (PMS) operates within the Soay sheep. Both positive and negative associations were identified between specific haplotypes and August weight, which is positively related to fitness (Milner *et al.* 1999; Jones *et al.* 2005), suggests that there may be fitness consequences to carrying particular haplotypes. The associations between immune measures and haplotypes also lend weight to PMS as the immunoglobulins assayed were anti-*Teladorsagia circumcincta*, which links the haplotypes to a nematode-related immune response. Finally, the association between haplotype C and FEC in adult males provides a direct link between the class IIa haplotypes and a pathogen measure. This supports the conclusions drawn in earlier work by Paterson *et al.* (1998) and Paterson (1998), although the specific allele associations with FEC were not fully replicated using haplotypes. A previously unidentified observation noted in Chapter 5 is the greater number of associations between haplotypes and traits in adults compared to lambs. Paterson *et al.* (1998) was only able to analyse *DRB1* associations in lambs and yearlings, and the vast majority of MHC studies in wild populations use juveniles individuals, perhaps because they typically represent the largest proportion of individuals in a single age class, or studies do not consider age at all (Appendix A). The analyses I have carried out in Chapter 5 clearly show that associations vary with age, and it may therefore be necessary to consider age-related effects in MHC studies.

The analysis of associations between haplotypes and traits within an animal model framework in Chapter 5 is novel for studies of the MHC in wild systems and makes the identified associations highly conservative. The animal model framework enabled me to account for relatedness between individuals by estimating the additive genetic component of each trait. This allowed the variation in the traits due to additive genetic

effects across the genome to be separated from variation that was associated with the class IIa haplotypes. The inclusion of this additive genetic effect removed a number of effects observed in the standard GLMMs, particularly a number of associations with immunoglobulin measures (see Chapter 5). This highlights the increased risk of detecting spurious results by not accounting for relatedness amongst sampled individuals in studies that focus on a single population.

Although it has not been possible to fully disentangle the evolutionary mechanisms acting within PMS within the scope of this thesis, I have been able to show that heterozygote advantage (using the typical overdominant meaning that heterozygous diplotypes are fitter than homozygotes diplotypes) is unlikely to play a major role in maintaining diversity within the class IIa region. There was no evidence of heterozygote excess at any of the LH stages or within the standing population of any given year. No associations between phenotypes and heterozygosity were identified as either a homozygous/heterozygous binary measure over all haplotypes nor as heterozygosity involving specific haplotypes. Lack of evidence for heterozygote advantage and associations between haplotypes and phenotypes leads to the expectation that negative frequency-dependent selection (NFDS) and/or fluctuating selection (FS) are operating in the Soay sheep (Piertney & Oliver 2006; Spurgin & Richardson 2010).

## 6.5 Future work

### 6.5.1 Fitness and other parasite measures

Although links between class IIa haplotypes and phenotypes were clearly identified in Chapter 5, there was a fairly bewildering array of associations, and so to truly understand whether these haplotypes are under selection, it will be necessary to investigate how they relate to fitness. The links between MHC and parasite burden (as measured by FEC), life-history traits (e.g. growth approximated via weight) and immune function (measured here by immunoglobulin measures) assessed in Chapter 5 do not directly address fitness, and may not reflect selection on or via MHC class IIa haplotypes. Lambs are under strongest selection for resistance to GIN (Hayward *et al.* 2011) and lamb and yearling survival is typically lower than in adults (Clutton-Brock *et al.* 1992), so first-year and second-year over-winter survival may reveal interesting relationships with haplotypes. Additionally, associations with over-winter survival and

annual-breeding success in adults as measures of annual fitness and with lifetime reproductive success as a measure of lifetime fitness would help to understand how haplotypes relate to fitness, and ultimately selection.

Within chapter 5, I only analysed associations between class IIa haplotypes and FEC, which only includes strongyle-type eggs. Due to time constraints, I was unable to consider other parasite measures which have been collected for the Soay sheep such as coccidia faecal oocyst count (FOC). FOC has been shown to be heritable in lambs ($h^2$ = 0.22 ± 0.21), though heritability was low in adults ($h^2$ = 0.06 ± 0.03) (Beraldi *et al.* 2007); class IIa haplotypes could therefore form part of the genetic basis for resistance to FOC. Additionally, the tapeworm *Moneizia expansa* has been assessed as present or absent in faeces throughout the study period but occurs at low prevalence (Craig *et al.* 2008), and ectoparasitic keds (*Melophagus ovinus*) have been routinely counted on captured sheep, though heritability of ked counts is minimal (Beraldi *et al.* 2007) and they are prevalent primarily in lambs (Craig *et al.* 2008). GIN, as measured by FEC, are thought to be the most important (measured) component of the pathogen community in terms of fitness in the Soay sheep, and therefore likely to have the strongest associations with class IIa haplotypes. However, selection on haplotypes may differ between pathogens and so it would be valuable to test alternative pathogens such as FOC, keds and *Moneizia*, and perhaps to do so in a multivariate analysis including FEC, such as co-inertia where a genetic correspondence analysis is combined with a parasitological principal components analysis (e.g. Tollenaere *et al.* 2008; Cutrera *et al.* 2011; Froeschke & Sommer 2012; Seifertová *et al.* 2016).

### 6.5.2 Negative frequency-dependent selection and fluctuating selection

Whilst Chapter 5 has demonstrated that heterozygote advantage is unlikely to be a major driver of balancing selection at the MHC class IIa within the Soay sheep, it was difficult to assess either negative frequency-dependent selection or fluctuating selection. How, exactly, these hypotheses could be addressed is unclear, and there are no single-population examples in the literature which address them well, although some studies have detected associations between a rare-allele and reduced parasite burden or fitness (Schad, Ganzhorn & Sommer 2005; Bolnick & Stutz 2017), or a change in frequency of a particular allele from rare to common (Codner *et al.* 2012). Additionally, either or both negative frequency-dependent selection or fluctuating

selection mechanisms could be operating at the same time. Allele frequency change may not happen rapidly, but it would be valuable to assess whether there is any temporal variation in class IIa haplotype frequencies since 1985 in the Soay sheep, whilst controlling for genetic drift by comparing against neutral markers. Using the same OLADRB microsatellite as Paterson *et al.* (1998), Charbonnel and Pemberton (2005) found some evidence for temporal variation in DRB1 genotypes but only within the eastern-most subunit (heft) of the population. More in depth understanding of how or whether gastrointestinal nematodes vary within the population, at both the species (e.g. individual variation in peptides that can be presented by MHC class IIa molecules) or community level (e.g. variation in the species compositions), and both spatially and temporally, may help to shed some light on whether there could be fluctuating selection imposed on the Soay sheep MHC.

### 6.5.3 Haplotype functional variation

Throughout this thesis, I have made the assumption that the class IIa haplotypes are equally different. However, this may not be true for three reasons. Firstly, there is some degree of allele sharing among some haplotypes (see Chapter 2), which means that some haplotypes are more similar to one another than to other haplotypes. Secondly, variation at the peptide binding region (PBR) is probably most important in determining the array of peptides that can be presented, yet how different the PBRs were of different alleles was not assessed. Thirdly, selection can only be imposed by the pathogens experienced by individuals, therefore while a diplotype may be capable of presenting a wide array of peptides, the selective advantage is only present in the presence of those pathogens.

The data presented in Chapter 5 demonstrates that it is unlikely that heterozygote advantage is operating in the Soay sheep under the assumption that haplotypes were equally different. However, I have been unable to exclude it entirely as I was unable to assess the functional variation among the haplotypes and test the divergent allele hypothesis (Wakeland *et al.* 1990). Heterozygote advantage proposes that heterozygotes are able to present a wider array of peptides than homozygotes, and therefore have greater fitness because they can mount immune responses to a wider range of pathogens. It may not be true, however, that all haplotypes present equally different sets of peptides, and many peptides may be presented by multiple haplotypes (Stoffels & Spencer 2008). Therefore, it may be that individuals carrying divergent

haplotypes actually have a fitness advantage compared to individuals carrying two haplotypes which present a similar array of haplotypes. Recent theoretical work by (Stefan 2016) has shown that divergent allele advantage is capable of maintaining the levels of variation observed at the MHC, and it would be interesting to test this hypothesis in the Soay sheep.

Determining the functional divergence of MHC alleles and haplotypes is not a simple matter. Studies of human HLA molecules have shown that alleles with similar amino acids in the peptide binding region (PBR) are able to bind similar peptide motifs (note that not all amino acids in the peptide are involved in binding to the MHC molecule) (del Guercio *et al.* 1995; Sidney *et al.* 1996; Sette & Sidney 1998, 1999; Castelli *et al.* 2002). This overlap in the ability of different alleles to present highly similar peptides has led to the classification of functionally similar HLA alleles into supertypes (Ou, Mitchell & Tingle 1998; Baas, Gao & Chelvanayagam 1999; Trachtenberg *et al.* 2003; Lund *et al.* 2004; Doytchinova & Flower 2005). Peptide binding assays are technically demanding and impractical on a large scale, so supertypes have been defined on the basis of the physio-chemical properties of the amino acids, by creating a matrix of the z-descriptors ($z_1$ (hydrophobicity), $z_2$ (steric bulk), $z_3$ (polarity), $z_4$, and $z_5$ (electronic effects) (Sandberg *et al.* 1998)) and clustering using multivariate analysis (Doytchinova & Flower 2005). This method has been successful in wild systems, particularly those with highly variable class I genes and when multiple loci co-amplify (e.g. primates – Schwensow *et al.* 2007; Huchard *et al.* 2008, 2010; birds – Sepil *et al.* 2012, 2013a; Radwan *et al.* 2012; Jones *et al.* 2014; amphibians – Teacher *et al.* 2009; Lillie *et al.* 2015; Tracy *et al.* 2015).

Being able to functionally classify the class IIa haplotypes of the Soay sheep would enable analyses comparing the level of divergence of the two haplotypes. In order to do so, two things must be achieved. Firstly, the nucleotide sequences of exon 2 of *DQB1* and *DQB2* on haplotype D need to obtained, and secondly, the amino acids involved in the peptide binding regions need to be determined. Sequencing of the haplotype D *DQB* alleles may be simplified with the identification of six haplotype D homozygotes from KASP genotyping, and would enable a simplistic calculation of amino acid divergence between haplotypes. The peptide binding regions of alleles are typically assumed to be the same as those identified from human HLA class II alleles (Reche & Reinherz 2003), which may or may not be accurate. Thus, it would be prudent to identify the positively selected sites within exon 2 of each locus (e.g. Tracy *et al.* 2015; Pearson *et al.* 2017).

### 6.5.4 Temporal variation in MHC diversity

Under both NFDS and FS, temporal variation in MHC diversity is predicted (Spurgin & Richardson 2010), either due to shifts from rare to common under NFDS or due to temporal and spatial variation in selection pressures under FS. Intensive sampling of MHC genotypes or diplotypes over multiple years within a wild population is rare and thus the data generated here for the Soay sheep represent a fairly unique opportunity to test for NFDS and FS. Nevertheless, it is uncertain whether the time span of the data (27 years of MHC haplotypes) is sufficient for such evolutionary processes to act, particularly in a species with a generation length of between four and five years (Coulson, Tuljapurkar & Childs 2010). Charbonnel and Pemberton (2005) showed that there may be temporal variation in diversity at the *DRB1* microsatellite in one of three spatially defined areas of the study population, hefts, between 1985 and 2000 by assessing the measure of population subdivision $F_{ST}$ in comparison with neutral microsatellites. However, this method is somewhat unsatisfactory as the three hefts defined by Coulson *et al.* (1999) are not geographically defined, movement between the hefts is frequent and the stability of the hefts over time is unknown. The few studies that have assessed temporal variation at the MHC in other species have typically avoided the problem of autocorrelation presented by alleles in generation *t+1* being non-independent from the alleles present in generation *t* by comparing diversity in populations at two distant time points (DeFaveri & Merilä 2015; Tison *et al.* 2015; Ciborowski *et al.* 2017) or within species with short generation times (Osborne *et al.* 2017). How exactly to address the question of whether class IIa haplotypes frequencies vary temporally in an intensively sampled pedigree is uncertain but possibilities include coalescent simulations of allele frequencies (Pečnerová *et al.* 2016) and gene-dropping simulations in which observed allele frequencies are compared to simulated allele frequencies under neutrality across the actual pedigree (as in Gratten *et al.* 2012; Grueber *et al.* 2012; Johnston *et al.* 2013).

### 6.5.5 Accurate mapping of class II a genes

The exact positions of class IIa loci are poorly understood, and the locations of *DQA2-like* and *DQB2-like* genes remain unknown. Development of SNP panels and resequencing efforts for the MHC region would benefit from improved mapping of this region, as would questions about how particular loci, haplotypes and variation in gene content affect fitness, whether evolution maintains particular alleles on haplotypes and

the contribution of gene duplication in this region. Highly similar alleles at different loci and potential gene variation between haplotypes make contig assembly during whole genome and resequencing efforts difficult when using NGS technologies, which typically result in short read lengths between 100 and 300 bp. The sheep genome was assembled using Illumina technologies with 100 – 200bp read lengths (Jiang *et al.* 2014), and though this assembly was facilitated by a variety of methods including SNP linkage maps, the MHC region on this assembly consists of a single haplotype structure that is actually a composite of multiple individuals. Highly polymorphic gene regions are notably difficult to assemble and therefore poorly represented by whole genome sequencing efforts (Ekblom & Wolf 2014). This is exacerbated when sequencing individuals which are heterozygous across this region, as it becomes increasingly difficult to assemble contigs across a region which differs in gene structure between the two chromosomes. Thus far, sequence maps focused purely on the MHC regions have been attempted using bacterial artificial chromosome clones and Sanger shotgun sequencing from a single Chinese merino sheep for class II (Gao *et al.* 2010) and a single Texel sheep for class I (Siva Subramaniam *et al.* 2015). PacBio sequencing, a relatively recent development in NGS technology, is capable of achieving read lengths of 10kb – 60kb (Rhoads & Au 2015). MHC targeted resequencing efforts using PacBio may therefore prove valuable in determining the genomic structure of the MHC region on different haplotypes.

### 6.5.6 Class I MHC variation

Throughout this thesis I focused only on class IIa variation because of existing SBG methods and the clear link between class IIa and presentation of peptides from macroparasites, such as GIN, which are relatively well understood in the Soay sheep (see Chapter 1). Class I genes in sheep are less well characterised than class IIa (but see Miltiadou *et al.* 2005; Ballingall *et al.* 2008), and are primarily linked to intracellular pathogens, which are less prevalent in Soay sheep (Graham *et al.* 2016) and therefore their effects on fitness are likely to be less pronounced than macroparasites. Class I gene content is typically more variable in number than in class II, and existing primers co-amplify all loci from cDNA (K. T. Ballingall, pers. comm.), making locus-specific genotypes unattainable. In a preliminary analysis not described here, I used NGS (MiSeq) to sequence the exons 2 and 3 of class I loci from 25 class IIa homozygous Soay sheep and found that the class I variation was much greater than at class IIa and did not correspond especially well to the class IIa haplotypes (data available from J. Pemberton

on request). More extensive class I sequencing could yield interesting insights into the variation in this region, given the relatively limited intracellular pathogens in Soay sheep compared to mainland domestic sheep (Graham *et al.* 2016) but it will be a large task.

## 6.6 Concluding remarks

This thesis has characterised the MHC class IIa haplotypes, developed a rapid haplotyping method, and generated a large diplotype data set for the Soay sheep of St. Kilda. This thesis highlights the challenges in generating high-quality, locus-specific genotype data for a non-model organism. It also highlights the value in characterising the haplotypes, as no single locus in the class IIa region was able to represent the full haplotypic variation observed in the Soay sheep. The evidence of balancing selection and associations between class IIa haplotypes and phenotypes related to health offer a promising glimpse into the evolutionary mechanisms which may be operating to maintain diversity within this region. This large MHC class IIa data set will hopefully be beneficial in disentangling PMS in the Soay sheep.

# References


Ackerman, A.L. & Cresswell, P. (2004) Cellular mechanisms governing cross-presentation of exogenous antigens. *Nature Immunology*, **5**, 678–684.

Acosta-Rodríguez, R., Alonso-Morales, R., Balladares, S., Flores-Aguilar, H., García-Vazquez, Z. & Gorodezky, C. (2005) Analysis of BoLA class II microsatellites in cattle infested with Boophilus microplus ticks: class II is probably associated with susceptibility. *Veterinary Parasitology*, **127**, 313–321.

Aeschlimann, P.B., Häberli, M.A., Reusch, T.B.H., Boehm, T. & Milinski, M. (2003) Female sticklebacks Gasterosteus aculeatus use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology*, **54**, 119–126.

Aguilar, A. & Garza, J.C. (2006) A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal steelhead (Oncorhynchus mykiss Walbaum). *Molecular Ecology*, **15**, 923–937.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D. & Wayne, R.K. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *PNAS*, **101**, 3490–3494.

Alcaide, M., Edwards, S. V. & Negro, J.J. (2007) Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *Journal of Molecular Evolution*, **65**, 541–554.

Alcaide, M., Edwards, S. V., Negro, J.J., Serrano, D. & Tella, J.L. (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (Falco naumanni). *Molecular Ecology*, **17**, 2652–2665.

Alcaide, M., Rodríguez, A., Negro, J.J. & Serrano, D. (2012) Male transmission ratio distortion supports MHC-linked cryptic female choice in the lesser kestrel (Aves: Falconidae). *Behavioral Ecology and Sociobiology*, **66**, 1467–1473.

Ali, A.O.A., Stear, A., Fairlie-Clarke, K., Brujeni, G.N., Isa, N.M.M., Salisi, M.S. Bin, Donskow-Łysoniewska, K., Groth, D., Buitkamp, J. & Stear, M.J. (2016) The genetic architecture of the MHC class II region in British Texel sheep. *Immunogenetics*, 1–7.

Allison, A.C. (1956) The sickle-cell and haemoglobin C genes in some African populations. *Annals of Human Genetics*, **21**, 67–89.

Amills, M., Ramiya, V., Norimine, J. & Lewin, H.A. (1998) The major histocompatibility complex of ruminants. *Revue scientifique et technique (International Office of Epizootics)*, **17**, 108–120.

Andersson, L., Gustafsson, K., Jonsson, A.K. & Rask, L. (1991) Concerted evolution in a segment of the first domain exon of polymorphic MHC class II β loci. *Immunogenetics*, **33**, 235–242.

Andersson, L., Lundén, A., Sigurdardottir, S., Davies, C.J. & Rask, L. (1988) Linkage relationships in the bovine MHC region. High recombination frequency between

class II subregions. *Immunogenetics*, **27**, 273–280.

Apanius, V., Penn, D., Slev, P.R., Ruff, L.R. & Potts, W.K. (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179–224.

Arbanasić, H., Duras, M., Podnar, M., Gomerčić, T., Ćurković, S. & Galov, A. (2014) Major histocompatibility complex class II variation in bottlenose dolphin from Adriatic Sea: inferences about the extent of balancing selection. *Marine Biology*, **161**, 2407–2422.

Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D. & Hedrick, P.W. (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (Oncorhynchus tshawytscha): effects of inbreeding and major histocompatibility complex genotypes. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 966–975.

Axtner, J. & Sommer, S. (2007) Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, Clethrionomys glareolus. *Immunogenetics*, **59**, 417–426.

Baas, A., Gao, X. & Chelvanayagam, G. (1999) Peptide binding motifs and specificities for HLA-DQ molecules. *Immunogenetics*, **50**, 8–15.

Babik, W. (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, **10**, 237–251.

Babik, W., Durka, W. & Radwan, J. (2005) Sequence diversity of the MHC DRB gene in the Eurasian beaver (Castor fiber). *Molecular Ecology*, **14**, 4249–4257.

Babik, W., Taberlet, P., Ejsmond, M.J. & Radwan, J. (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, **9**, 713–719.

de Bakker, P.I.W., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M., Morrison, J., Richardson, A., Walsh, E.C., Gao, X., Galver, L., Hart, J., Hafler, D.A., Pericak-Vance, M., Todd, J.A., Daly, M.J., Trowsdale, J., Wijmenga, C., Vyse, T.J., Beck, S., Murray, S.S., Carrington, M., Gregory, S., Deloukas, P. & Rioux, J.D. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genetics*, **38**, 1166–1172.

de Bakker, P.I.W. & Raychaudhuri, S. (2012) Interrogating the major histocompatibility complex with high-throughput genomics. *Human Molecular Genetics*, **21**, R29–R36.

Ballingall, K.T., Ellis, S.A., MacHugh, N.D., Archibald, S.D. & McKeever, D.J. (2004) The DY genes of the cattle MHC: Expression and comparative analysis of an unusual class II MHC gene pair. *Immunogenetics*, **55**, 748–755.

Ballingall, K.T., Lantier, I., Todd, H., Lantier, F. & Rocchi, M. (2017) Structural and functional diversity arising from intra and inter haplotype combinations of duplicated DQA and B loci within the Ovine MHC. *Immunogenetics*.

Ballingall, K.T., MacHugh, N.D., Taracha, E.L.N., Mertens, B., McKeever, D.J., Fakher, M., Wu, J., Qin, D., Szakal, A.K. & Tew, J.G. (2001) Transcription of the unique ruminant class II major histocompatibility complex-DYA and DIB genes in dendritic cells. *European Journal of Immunology*, **31**, 82–86.

Ballingall, K.T. & McKeever, D.J. (2005) Conservation of promoter, coding and intronic

regions of the nonclassical MHC class II DYA gene suggests evolution under functional constraints. *Animal Genetics*, **36**, 237–239.

Ballingall, K.T., Miltiadou, D., Chai, Z.W., Mclean, K., Rocchi, M., Yaga, R. & McKeever, D.J. (2008) Genetic and proteomic analysis of the MHC class I repertoire from four ovine haplotypes. *Immunogenetics*, **60**, 177–184.

Ballingall, K.T., Rocchi, M., McKeever, D.J. & Wright, F. (2010) Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep. *PLoS ONE*, **5**, e11402.

Ballingall, K.T., Steele, P., Lantier, I., Cotelli, M., Todd, H., Lopez, G., Martin, E. & Lantier, F. (2015) An ancient interlocus recombination increases class II MHC DQA diversity in sheep and other Bovidae. *Animal Genetics*, **46**, 333–336.

Ballingall, K.T. & Tassi, R. (2010) Sequence-based genotyping of the sheep MHC class II DRB1 locus. *Immunogenetics*, **62**, 31–39.

Ballingall, K.T., Wright, H., Redmond, J., Dutia, B.M., Hopkins, J., Lang, J., Deverson, E. V, Howard, J.C., Puri, N. & Haig, D. (1992) Expression and characterization of ovine major histocompatibility complex class II (OLA-DR) genes. *Animal Genetics*, **23**, 347–359.

Banks, S.C., Dubach, J., Viggers, K.L. & Lindenmayer, D.B. (2010) Adult survival and microsatellite diversity in possums: Effects of major histocompatibility complex-linked microsatellite diversity but not multilocus inbreeding estimators. *Oecologia*, **162**, 359–370.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Basha, G., Omilusik, K., Chavez-Steenbock, A., Reinicke, A.T., Lack, N., Choi, K.B. & Jefferies, W.A. (2012) A CD74-dependent MHC class I endolysosomal cross-presentation pathway. *Nature Immunology*, **13**, 237–245.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1–48.

Bateson, Z.W., Hammerly, S.C., Johnson, J.A., Morrow, M.E., Whittingham, L.A. & Dunn, P.O. (2016) Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, **25**, 4730–4744.

Belich, M.P., Madrigal, J.A., Hildebrand, W.H., Zemmour, J., Williams, R.C., Luz, R., Petzl-Erler, M.L. & Parham, P. (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature*, **357**, 326–329.

Beraldi, D., McRae, A.F., Gratten, J., Pilkington, J.G., Slate, J., Visscher, P.M. & Pemberton, J.M. (2007) Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (Ovis aries). *International Journal for Parasitology*, **37**, 121–129.

Bérénos, C., Ellis, P. a., Pilkington, J.G., Lee, S.H., Gratten, J. & Pemberton, J.M. (2015) Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, **24**, 1810–1830.

Bérénos, C., Ellis, P.A., Pilkington, J.G. & Pemberton, J.M. (2014) Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. *Molecular Ecology*, **23**, 3434–3451.

References

Bernatchez, L. & Landry, C. (2003) MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bichet, C., Moodley, Y., Penn, D.J., Sorci, G. & Garnier, S. (2015) Genetic structure in insular and mainland populations of house sparrows (Passer domesticus) and their hemosporidian parasites. *Ecology and Evolution*, **5**, 1639–1652.

Biedrzycka, A., Kloch, A., Buczek, M. & Radwan, J. (2011) Major histocompatibility complex DRB genes and blood parasite loads in fragmented populations of the spotted suslik Spermophilus suslicus. *Mammalian Biology*, **76**, 672–677.

Black, F.L. & Hedrick, P.W. (1997) Strong balancing selection at HLA loci: evidence from segregation in South Amerindian families. *Proc Natl Acad Sci U S A*, **94**, 12452–12456.

Black, F.L. & Salzano, F.M. (1981) Evidence for heterosis in the HLA system. *American journal of human genetics*, **33**, 894–899.

Bodmer, W.. (1972) Evolutinary Significance of the HL-A System. *Nature*, **237**, 139–145.

De Boer, R.J., Borghans, J.A.M., Van Boven, M., Keşmir, C. & Weissing, F.J. (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics*, **55**, 725–731.

Bollmer, J.L., Dunn, P.O., Whittingham, L.A. & Wimpee, C. (2010) Extensive MHC class II B gene duplication in a passerine, the common yellowthroat (Geothlypis trichas). *Journal of Heredity*, **101**, 448–460.

Bollmer, J.L., Vargas, F.H. & Parker, P.G. (2007) Low MHC variation in the endangered Galapagos penguin (Spheniscus mendiculus). *Immunogenetics*, **59**, 593–602.

Bolnick, D.I. & Stutz, W.E. (2017) Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature*, **546**, 285–288.

Bonneaud, C., Mazuc, J., Chastel, O., Westerdahl, H. & Sorci, G. (2004) Terminal investment induced by immune challenge and fitness traits associated with major histocompatibility complex in the house sparrow. *Evolution*, **58**, 2823.

Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. & Sorci, G. (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution*, **60**, 383–389.

Bonneaud, C., Richard, M., Faivre, B., Westerdahl, H. & Sorci, G. (2005) An Mhc class I allele associated to the expression of T-dependent immune response in the house sparrow. *Immunogenetics*, **57**, 782–789.

Borghans, J.A.M., Noest, A.J. & De Boer, R.J. (2003) Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology*, **33**, 3353–3358.

Boyce, W.M., Hedrick, P.W., Muggli-Cockett, N.E., Kalinowski, S., Penedo, M.C.T. & Ramey, R.R. (1997) Genetic Variation of Major Histocompatibility Complex and Microsatellite Loci: A Comparison in Bighorn Sheep. *Genetics*, **145**, 421–433.

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

Boyd, H.E.G. (1999) *Early Development of Parasitism in Soay Sheep*. University of Cambridge.

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**, 129–164.

Brouwer, L., Barr, I., Van De Pol, M., Burke, T., Komdeur, J. & Richardson, D.S. (2010) MHC-dependent survival in a wild population: Evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, **19**, 3444–3455.

Brown, J.L. & Eklund, A. (1994) Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *The American Naturalist*, **143**, 435–461.

Browning, S.R. & Browning, B.L. (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, **81**, 1084–1097.

Bryja, J., Charbonnel, N., Berthier, K., Galan, M. & Cosson, J.F. (2007) Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Molecular Ecology*, **16**, 5084–5097.

Buitkamp, J. & Epplen, J.T. (1996) Major histocompatibility and T-cell receptor genes in artiodactyls: Characterization, polymorphism and genetic resistance to a helmintic infection. *Journal of Animal Breeding and Genetics*, **113**, 287–291.

Buitkamp, J., Filmether, P., Stear, M.J., Epplen, J.T., Buítkamp, J., Filmether, P., Stear, M.J. & Epplen, J.T. (1996) Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly Ostertagia circumcincta infection. *Parasitology research*, **82**, 693–696.

Van Den Bussche, R.A., Ross, T.G. & Hoofer, S.R. (2002) Genetic variation at a major histocompatibility locus within and among populations of white-tailed deer (Odocoileus virginianus). *Journal of Mammalogy*, **83**, 31–39.

Butler, J.M. (2005) *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*, 2nd ed. Elsevier, Burlington.

Cammen, K.M., Hoffman, J.I., Knapp, L.A., Harwood, J. & Amos, W. (2011) Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (Halichoerus grypus). *Molecular Ecology*, **20**, 740–752.

Campos, J.L., Bellocq, J.G. de, Schaschl, H. & Suchentrunk, F. (2011) MHC class II DQA gene variation across cohorts of brown hares (Lepus europaeus) from eastern Austria: Testing for different selection hypotheses. *Mammalian Biology - Zeitschrift für Säugetierkunde*, **76**, 251–257.

Campos, J.L., Posada, D. & Morán, P. (2006) Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (Salmo trutta). *Conservation Genetics*, **7**, 515–530.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K. & O'Brien, S.J. (1999) HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. *Science*, **283**, 1748–1752.

Castelli, F.A., Buhot, C., Sanson, A., Zarour, H., Pouvelle-Moratille, S., Nonn, C., Gahery-Segard, H., Guillet, J.-G., Menez, A., Georges, B. & Maillere, B. (2002) HLA-DP4, the Most Frequent HLA II Molecule, Defines a New Supertype of Peptide-Binding Specificity. *The Journal of Immunology*, **169**, 6928–6934.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

References

Charbonnel, N. & Pemberton, J.M. (2005) A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*, **95**, 377–388.

Chardon, P., Kirszenbaum, M., Cullen, P.R., Geffrotin, C., Auffray, C., Strominger, J.L., Cohen, D. & Vaiman, M. (1985) Analysis of the sheep MHC using HLA class I, II, and C4 cDNA probes. *Immunogenetics*, **22**, 349–358.

Charlesworth, B. (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Charlesworth, D. & Awadalla, P. (1998) Flowering plant self-incompatibility: the molecular population genetics of Brassica S-loci. *Heredity*, **81**, 1–9.

Charlesworth, D., Awadalla, P., Mable, B.K. & Schierup, M.H. (2000) Population-level studies of multiallelic serf-incompatibility loci, with particular reference to Brassicaceae. *Annals of Botany*, **85**, 227–239.

Charlesworth, D. & Willis, J.H. (2009) The genetics of inbreeding depression. *Nature reviews. Genetics*, **10**, 783–96.

Charon, K., Moskwa, B., Rutkowski, R., Gruszczyńska, J. & Świderek, W. (2002) Microsatellite polymorphism in DRB1 gene (MHC class II) and its relation to nematode faecal egg count in Polish Heath Sheep. *Journal of Animal and Feed Sciences*, **11**, 47–58.

Chaves, L.D., Faile, G.M., Krueth, S.B., Hendrickson, J.A. & Reed, K.M. (2010) Haplotype variation, recombination, and gene conversion within the turkey MHC-B locus. *Immunogenetics*, **62**, 465–477.

Chen, J.J., Hollenbach, J.A., Trachtenberg, E., Just, J.J., Carrington, M., Rønningen, K.S., Begovich, A., King, M.-C., McWeeney, S., Mack, S.J., Erlich, H.A. & Thomson, G. (1999) Hardy-Weinberg testing for HLA class II (DRB1, DQA1, DQB1, AND DPB1) loci in 26 human ethnic groups. *Tissue Antigens*, **54**, 533–542.

Ciborowski, K., Jordan, W.C., Garcia de Leaniz, C. & Consuegra, S. (2017) Temporal and spatial instability in neutral and adaptive (MHC) genetic variation in marginal salmon populations. *Scientific Reports*, **7**, 42416.

Clutton-Brock, T.H., Grenfell, B.T., Coulson, T., MacColl, A.D.C., Illius, A.W., Forchhammer, M.., Wilson, K., Lindstrom, J., Crawley, M.J. & Albon, S.D. (2004a) Population dynamics in Soay sheep. *Soay Sheep: Dynamics and selection in an island population* (eds T.H. Clutton-Brock & J.M. Pemberton), pp. 52–82. Cambridge University Press, Cambridge.

Clutton-Brock, T.H., Pemberton, J.M. & Coulson, T. (2004b) The sheep of St Kilda. *Soay Sheep: Dynamics and selection in an island population* (eds T.H. Clutton-Brock & J.M. Pemberton), pp. 17–50. Cambridge University Press, Cambridge.

Clutton-Brock, T.H., Price, O.F., Albon, S.D. & Jewell, P.A. (1992) Early Development and Population Fluctuations in Soay Sheep. *Journal of Animal Ecology*, **61**, 381–396.

Codner, G.F., Stear, M.J., Reeve, R., Matthews, L. & Ellis, S.A. (2012) Selective forces shaping diversity in the class I region of the major histocompatibility complex in dairy cattle. *Animal Genetics*, **43**, 239–249.

Collin, H., Burri, R., Comtesse, F. & Fumagalli, L. (2013) Combining molecular evolution and environmental genomics to unravel adaptive processes of MHC class IIB

diversity in european minnows (phoxinus phoxinus). *Ecology and Evolution*, **3**, 2568–2585.

Coltman, D.W., Pilkington, J.G., Kruuk, L.E.B., Wilson, K. & Pemberton, J.M. (2001a) Positive Genetic Correlation between Parasite Resistance and Body Size in a Free-Living Ungulate Population. *Evolution*, **55**, 2116–2125.

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

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.

Consuegra, S., de Eyto, E., McGinnity, P., Stet, R.J.M. & Jordan, W.C. (2011) Contrasting responses to selection in class I and class IIα major histocompatibility-linked markers in salmon. *Heredity*, **107**, 143–154.

Corbishley, A., Connelley, T.K., Wolfson, E.B., Ballingall, K.T., Beckett, A.E., Galle, D.L. & McNeilly, T.N. (2016) Identification of epitopes recognised by mucosal CD4+ T-cell populations from cattle experimentally colonised with Escherichia coli O157:H7. *Veterinary Research*, **47**, 90.

Cortes, A. & Brown, M.A. (2011) Promise and pitfalls of the Immunochip. *Arthritis Research & Therapy*, **13**, 101.

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.

Coulson, T., Tuljapurkar, S. & Childs, D.Z. (2010) Using evolutionary demography to link life history theory, quantitative genetics and population ecology. *Journal of Animal Ecology*, **79**, 1226–1240.

Craig, B.H., Jones, O.R., Pilkington, J.G. & Pemberton, J.M. (2009) Re-establishment of nematode infra-community and host survivorship in wild Soay sheep following anthelmintic treatment. *Veterinary Parasitology*, **161**, 47–52.

Craig, B.H., Pilkington, J.G. & Pemberton, J.M. (2006) Gastrointestinal nematode species burdens and host mortality in a feral sheep population. *Parasitology*, **133**, 485–496.

Craig, B.H., Tempest, L.J., Pilkington, J.G. & Pemberton, J.M. (2008) Metazoan-protozoan parasite co-infections and host body weight in St Kilda Soay sheep. *Parasitology*, **135**, 433–441.

Crawford, A., Dodds, K., Ede, A., Pierson, C., Montgomery, G., Garmonsway, H., Beattie, A., Davies, K., Maddox, J. & Kappes, S. (1995) An Autosomal Genetic Linkage Map of the Sheep Genome. *Genetics*, **140**, 703–724.

Crawley, M.J., Albon, S.D., Bazely, D.. & Milner, J.M. (2004) Vegetation and sheep population dynamics. *Soay Sheep: Dynamics and selection in an island population* (eds T.H. Clutton-Brock & J.M. Pemberton), pp. 97–105. Cambridge University Press, Cambridge.

References

Croisetière, S., Tarte, P.D., Bernatchez, L. & Belhumeur, P. (2008) Identification of MHC class IIβ resistance/susceptibility alleles to Aeromonas salmonicida in brook charr (Salvelinus fontinalis). *Molecular Immunology*, **45**, 3107–3116.

Cullen, M., Noble, J., Erlich, H., Thorpe, K., Beck, S., Klitz, W., Trowsdale, J. & Carrington, M. (1997) Characterization of recombination in the HLA class II region. *American journal of human genetics*, **60**, 397–407.

Cutrera, A.P., Zenuto, R.R. & Lacey, E.A. (2011) MHC variation, multiple simultaneous infections and physiological condition in the subterranean rodent Ctenomys talarum. *Infection, Genetics and Evolution*, **11**, 1023–1036.

Danchin, E., Vitiello, V., Vienne, A., Richard, O., Gouret, P., McDermott, M.F. & Pontarotti, P. (2004) The major histocompatibility complex origin. *Immunol Rev*, **198**, 216–232.

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*, **96**, 252–258.

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P. & Kulski, J.K. (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunological Reviews*, **167**, 275–304.

DeFaveri, J. & Merilä, J. (2015) Temporal stability of genetic variability and differentiation in the three-spined stickleback ( Gasterosteus aculeatus). *PLoS ONE*, **10**, e0123891.

Dehkordi, H.S. & Zamani, P. (2015) New mutations in DQA2 exon 2 and their associations with some production traits and serum parameters of sheep. *Small Ruminant Research*, **130**, 47–53.

Deter, J., Bryja, J., Chaval, Y., Galan, M., Henttonen, H., Laakkonen, J., Voutilainen, L., Vapalahti, O., Vaheri, A., Salvador, A.R., Morand, S., Cosson, J.-F. & Charbonnel, N. (2008) Association between the DQA MHC class II gene and Puumala virus infection in Myodes glareolus, the bank vole. *Infection, Genetics and Evolution*, **8**, 450–458.

Deverson, E. V, Wright, H., Watson, S., Ballingall, K.T., Huskisson, N., Diamond, A.G. & Howard, J.C. (1991) Class II major histocompatibility complex genes of the sheep. *Animal Genetics*, **22**, 211–225.

Dilthey, A.T., Moutsianas, L., Leslie, S. & McVean, G. (2011) HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics*, **27**, 968–972.

Dionne, M., Miller, K.M., Dodson, J.J. & Bernatchez, L. (2009) MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **364**, 1555–1565.

Dionne, M., Miller, K.M., Dodson, J.J., Caron, F. & Bernatchez, L. (2007) Clinal variation in MHC diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, **61**, 2154–2164.

Ditchkoff, S.S., Hoofer, S.R., Lochmiller, R.L., Masters, R.E. & Van Den Bussche, R.A. (2005) *MHC-DRB* evolution provides insight into parasite resistance in white-tailed deer (ed CW Edwards). *The Southwestern Naturalist*, **50**, 57–64.

Doherty, P.C. & Zinkernagel, R.M. (1975) Enhanced immunological surveillance in mice

heterozygous at the H-2 gene complex. *Nature*, **256**, 50–52.

Doytchinova, I.A. & Flower, D.R. (2005) In silico identification of supertypes for class II MHCs. *Journal of Immunology*, **174**, 7085–7095.

Dukkipati, V.S.R., Blair, H.T., Garrigan, D. & Murray, A. (2006a) "Ovar-Mhc" - Ovine major histocompatibility complex: Structure and gene polymorphisms. *Genetics and Molecular Research*, **5**, 581–608.

Dukkipati, V.S.R., Blair, H.T., Garrigan, D. & Murray, A. (2006b) Ovar-Mhc--ovine major histocompatibility complex: role in genetic resistance to diseases. *New Zealand Veterinary Journal*, **54**, 153–160.

Dunn, P.O., Bollmer, J.L., Freeman-Gallant, C.R. & Whittingham, L.A. (2013) Mhc variation is related to a sexually selected ornament, Survival, and parasite resistance in common yellowthroats. *Evolution*, **67**, 679–687.

Edwards, S. V., Gasper, J., Garrigan, D., Martindale, D. & Koop, B.F. (2000) A 39-kb Sequence Around a Blackbird Mhc Class II Gene: Ghost of Selection Past and Songbird Genome Architecture. *Molecular Biology and Evolution*, **17**, 1384–1395.

Edwards, S. V., Hess, C.M., Gasper, J. & Garrigan, D. (1999) Toward an evolutionary genomics of the avian Mhc. *Immunological Reviews*, **167**, 119–132.

van Eijk, M.J.T., Beever, J.E., Da, Y., Stewart, J.A., Nicholaides, G.E., Green, C.A. & Lewin, H.A. (1995) Genetic mapping of BoLA-A, CYP21, DRB3, DYA, and PRL on BTA23. *Mammalian Genome*, **6**, 151–154.

Eimes, J.A., Townsend, A.K., Sepil, I., Nishiumi, I. & Satta, Y. (2015) Patterns of evolution of MHC class II genes of crows (Corvus) suggest trans-species polymorphism. *PeerJ*, **3**, e853.

Eizaguirre, C., Lenz, T.L. & Eklund, A. (2010) Major histocompatability complex polymorphism: Dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, **77**, 2023–2047.

Eizaguirre, C., Lenz, T.L., Kalbe, M. & Milinski, M. (2012a) Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, **15**, 723–731.

Eizaguirre, C., Lenz, T.L., Kalbe, M. & Milinski, M. (2012b) Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications*, **3**, 621.

Eizaguirre, C., Lenz, T.L., Sommerfeld, R.D., Harrod, C., Kalbe, M. & Milinski, M. (2011) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology*, **25**, 605–622.

Ekblom, R., Sæther, S.A., Jacobsson, P., Fiske, P., Sahlman, T., Grahn, M., Kålås, J.A. & Höglund, J. (2007) Spatial pattern of MHC class II variation in the great snipe (Gallinago media). *Molecular Ecology*, **16**, 1439–1451.

Ekblom, R. & Wolf, J.B.W. (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**, 1026–1042.

Elbers, J.P., Clostio, R.W. & Taylor, S.S. (2017) Neutral Genetic Processes Influence MHC Evolution in Threatened Gopher Tortoises (Gopherus polyphemus). *Journal of Heredity*, **108**, 515–523.

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature*

*reviews. Genetics*, **5**, 435–445.

Ellis, S.A., Holmes, E.C., Staines, K.A., Smith, K.B., Stear, M.J., McKeever, D.J., MacHugh, N.D. & Morrison, W.I. (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. *Immunogenetics*, **50**, 319–328.

Ellison, A., Allainguillaume, J., Girdwood, S., Pachebat, J., Peat, K., Wright, P. & Consuegra, S. (2012) Maintaining functional MHC diversity under inbreeding: The case of a selfing vertebrate. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 5004–13.

Ennen, S., Hamann, H., Distl, O., Hickford, J.G.H., Zhou, H. & Ganter, M. (2009) A field trial to control ovine footrot via vaccination and genetic markers. *Small Ruminant Research*, **86**, 22–25.

Escayg, A.., Hickford, J.G.H. & Bullock, D.. (1997) Association between alleles of the ovine major histocompatibility complex and resistance to footrot. *Research in Veterinary Science*, 283–287.

Evans, M.L., Neff, B.D. & Heath, D.D. (2010a) MHC-mediated local adaptation in reciprocally translocated Chinook salmon. *Conservation Genetics*, **11**, 2333–2342.

Evans, M.L., Neff, B.D. & Heath, D.D. (2010b) MHC genetic structure and divergence across populations of Chinook salmon (Oncorhynchus tshawytscha). *Heredity*, **104**, 449–459.

Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical population biology*, **3**, 87–112.

Fabb, S.A., Maddox, J.F., Gogolin-Ewens, K., Baker, L., Wu, M.-J. & RBrandon, M. (1993) Isolation, characterization and evolution of ovine major histocompatibility complex class II DRA and DQA genes. *Animal Genetics*, **24**, 249–255.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368–376.

Figueroa Castillo, J.A., Medina, R.D.M., Villalobos, J.M.B., Gayosso-Vázquez, A., Ulloa-Arvízu, R., Rodríguez, R.A., Ramírez, H.P. & Alonso Morales, R.A. (2011) Association between major histocompatibility complex microsatellites, fecal egg count, blood packed cell volume and blood eosinophilia in Pelibuey sheep infected with Haemonchus contortus. *Veterinary Parasitology*, **177**, 339–344.

Figueroa, F., Günther, E. & Klein, J. (1988a) MHC polymorphism pre-dating speciation. *Nature*, **335**, 265–267.

Figueroa, F., Günther, E. & Klein, J. (1988b) MHC polymorphism pre-dating speciation. *Nature*, **335**, 265–267.

Fijarczyk, A. & Babik, W. (2015) Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, n/a-n/a.

Flajnik, M.F. & Kasahara, M. (2010) Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nature Reviews Genetics*, **11**, 47–59.

Forrest, R.H.J., Zhou, H., Fang, Q., Smyth, A., Frampton, C.M. & Hickford, J.G.H. (2010) No evidence for a universal association between variation in faecal egg count for a mixed field-challenge of gastrointestinal parasites and the presence of the Ovar-DQA1 null haplotype in sheep. *Veterinary Immunology and Immunopathology*, **135**, 303–305.

Forsberg, L.A., Dannewitz, J., Petersson, E. & Grahn, M. (2007) Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout - females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology*, **20**, 1859–1869.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**, 233–249.

Fraser, B.A. & Neff, B.D. (2010) Parasite mediated homogenizing selection at the MHC in guppies. *Genetica*, **138**, 273–278.

Fraser, B.A., Ramnarine, I.W. & Neff, B.D. (2010) Selection at the MHC class IIB locus across guppy (Poecilia reticulata) populations. *Heredity*, **104**, 155–167.

Froeschke, G. & Sommer, S. (2005) MHC Class II DRB Variability and Parasite Load in the Striped Mouse (Rhabdomys pumilio) in the Southern Kalahari. *Molecular Biology and Evolution*, **22**, 1254–1259.

Froeschke, G. & Sommer, S. (2012) Insights into the complex associations between MHC class II DRB polymorphism and multiple gastrointestinal parasite infestations in the striped mouse. *PLoS ONE*, **7**, e31820.

Froeschke, G. & Sommer, S. (2014) Role of selection versus neutral processes determining genetic variation in a small mammal along a climatic gradient in southern Africa. *Evolutionary Ecology*, **28**, 1169–1190.

Gao, J., Liu, K., Liu, H., Blair, H.T., Li, G., Chen, C., Tan, P. & Ma, R.Z. (2010) A complete DNA sequence map of the ovine major histocompatibility complex. *BMC Genomics*, **11**, 466.

Garrigan, D. & Hedrick, P.W. (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics*, **53**, 483–489.

Garrigan, D. & Hedrick, P.W. (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.

Gelasakis, A.I., Arsenos, G., Hickford, J.G.H., Zhou, H., Psifidi, A., Valergakis, G.E. & Banos, G. (2013) Polymorphism of the MHC-DQA2 gene in the Chios dairy sheep population and its association with footrot. *Livestock Science*, **153**, 56–59.

Génin, E., Ober, C., Weitkamp, L. & Thomson, G. (2000) A robust test for assortative mating. *European Journal of Human Genetics*, **8**, 119–124.

Germain, R.N. (1986) Immunology: The ins and outs of antigen processing and presentations. *Nature*, **322**, 687–688.

Gill, P. (2001) Application of low copy number DNA profiling. *Croatian medical journal*, **42**, 229–232.

Gillet, R.M., Murray, B.W. & White, B.N. (2013) Characterization of Class I – and Class II – Like Major Histocompatibility Complex Loci in Pedigrees of North Atlantic Right Whales. *Journal of Heredity*, **105**, 188–202.

Gillingham, M.A.., Courtiol, A., Teixeira, M., Galan, M., Bechet, A. & Cezilly, F. (2016) Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species. *Journal of Evolutionary Biology*,

**29**, 438–454.

Gilmore, A., Gogel, B., Cullis, B. & Thompson, R. (2009) Asreml User Guide Release 3.0.

Go, Y., Satta, Y., Kawamoto, Y., Rakotoarisoa, G., Randrianjafy, A., Koyama, N. & Hirai, H. (2003) Frequent segmental sequence exchanges and rapid gene duplication characterize the MHC class I genes in lemurs. *Immunogenetics*, **55**, 450–461.

Godot, V., Harraga, S., Beurton, I., Tiberghien, P., Sarciron, E., Gottstein, B. & Vuitton, D.A. (2000) Resistance/susceptibility to Echinococcus multilocularis infection and cytokine profile in humans. II. Influence of the HLA B8, DR3, DQ2 haplotype. *Clinical and Experimental Immunology*, **121**, 491–498.

González-Quevedo, C., Davies, R.G., Phillips, K.P., Spurgin, L.G. & Richardson, D.S. (2016) Landscape-scale variation in an anthropogenic factor shapes immune gene variation within a wild population. *Molecular Ecology*, **25**, 4234–4246.

Graham, A.L., Hayward, A.D., Watt, K.A., Pilkington, J.G., Pemberton, J.M. & Nussey, D.H. (2010) Fitness correlates of heritable variation in antibody responsiveness in a wild mammal. *Science*, **330**, 662–665.

Graham, A.L., Nussey, D.H., Lloyd-Smith, J.O., Longbottom, D., Maley, M., Pemberton, J.M., Pilkington, J.G., Prager, K.C., Smith, L., Watt, K.A., Wilson, K., Mcneilly, T.N. & Brülisauer, F. (2016) Exposure to viral and bacterial pathogens among Soay sheep (Ovis aries) of the St Kilda archipelago. *Epidemiology and infection*, 1–10.

Gratten, J., Pilkington, J.G., Brown, E.A., Clutton-Brock, T.H., Pemberton, J.M. & Slate, J. (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology*, **21**, 2977–2990.

Grenfell, B.T., Wilson, K., Isham, V.S., Boyd, H.E.G. & Dietz, K. (1995) Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode–ruminant interactions as a case study. *Parasitology*, **111**, S135–S151.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S. & Stet, R.J.M. (2003) MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics*, **55**, 210–219.

Grueber, C.E., Wallis, G.P., King, T.M. & Jamieson, I.G. (2012) Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. *PLoS ONE*, **7**, e45011.

Gu, X. & Nei, M. (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol Biol Evol*, **16**, 147–156.

del Guercio, M.F., Sidney, J., Hermanson, G., Perez, C., Grey, H.M., Kubo, R.T. & Sette, A. (1995) Binding of a peptide antigen to multiple HLA alleles allows definition of an A2-like supertype. *Journal of Immunology*, **154**, 685–93.

Gulland, F.M.D. (1992) The role of nematode parasites in Soay sheep (Ovis aries L.) mortality during a population crash. *Parasitology*, **105 ( Pt 3**, 493–503.

Gulland, F.M.D., Albon, S.D., Pemberton, J.M., Moorcroft, P.R. & Clutton-Brock, T.H. (1993) Parasite-associated polymorphism in a cyclic ungulate population. *Proceedings of the Royal Society of London B: Biological Sciences*, **254**, 7–13.

Gulland, F.M.D. & Fox, M. (1992) Epidemiology of nematode infections of Soay sheep (Ovis aries L.) on St Kilda. *Parasitology*, **105**, 481–492.

Guo, S.W. & Thompson, E.A. (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, **48**, 361–372.

Gutierrez-Espeleta, G.A., Hedrick, P.W., Kalinowski, S.T., Garrigan, D. & Boyce, W.M. (2001) Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity*, **86**, 439–450.

Gutiérrez-Gil, B., Pérez, J., de la Fuente, L.F., Meana, A., Martínez-Valladares, M., San Primitivo, F., Rojo-Vázquez, F.A. & Arranz, J.J. (2010) Genetic parameters for resistance to trichostrongylid infection in dairy sheep. *Animal*, **4**, 505–512.

Hablützel, P.I., Vanhove, M.P.M., Grégoir, A.F., Hellemans, B., Volckaert, F.A.M. & Raeymaekers, J.A.M. (2014) Intermediate number of major histocompatibility complex class II *B* length variants relates to enlarged perivisceral fat deposits in the blunt-head cichlid *Tropheus moorii*. *Journal of Evolutionary Biology*, **27**, 2177–2190.

Hadfield, J.D., Richardson, D.S. & Burke, T. (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15**, 3715–3730.

Haldane, J.B.S. (1949) Disease and Evolution. *Current Science*, **63**, 599–604.

Haldane, J.B.S. (1954) An exact test for randomness of mating. *Journal of Genetics*, **52**, 631–635.

Hansen, M.M., Skaala, Ø., Jensen, L.F., Bekkevold, D. & Mensberg, K.-L.D. (2007) Gene flow, effective population size and selection at major histocompatibility complex genes: brown trout in the Hardanger Fjord, Norway. *Molecular Ecology*, **16**, 1413–1425.

Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P. & Du Pasquier, L. (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, Oncorhynchus mykiss. *Journal of Immunology*, **163**, 774–786.

Harf, R. & Sommer, S. (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, Gerbillurus paeba, in the southern Kalahari. *Molecular Ecology*, **14**, 85–91.

Hassan, M., Good, B., Hanrahan, J.P., Campion, D., Sayers, G., Mulcahy, G. & Sweeney, T. (2011) The dynamic influence of the DRB1*1101 allele on the resistance of sheep to experimental Teladorsagia circumcincta infection. *Veterinary Research*, **42**, 46.

Hawley, D.M. & Fleischer, R.C. (2012) Contrasting epidemic histories reveal pathogen-mediated balancing selection on class II MHC diversity in a wild songbird. *PLoS ONE*, **7**.

Hayward, A.D. (2013) Causes and consequences of intra- and inter-host heterogeneity in defence against nematodes. *Parasite Immunology*, **35**, 362–373.

Hayward, A.D., Garnier, R., Watt, K. a., Pilkington, J.G., Grenfell, B.T., Matthews, J.B., Pemberton, J.M., Nussey, D.H. & Graham, A.L. (2014a) Heritable, Heterogeneous, and Costly Resistance of Sheep against Nematodes and Potential Feedbacks to Epidemiological Dynamics. *The American Naturalist*, **184**, S58–S79.

Hayward, A.D., Nussey, D.H., Wilson, A.J., Berenos, C., Pilkington, J.G., Watt, K.A., Pemberton, J.M. & Graham, A.L. (2014b) Natural Selection on Individual Variation in Tolerance of Gastrointestinal Nematode Infection. *PLoS Biology*, **12**, 1–13.

Hayward, A.D., Wilson, A.J., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. &

Kruuk, L.E.B. (2011) Natural selection on a measure of parasite resistance varies across ages and environmental conditions in a wild mammal. *Journal of Evolutionary Biology*, **24**, 1664–1676.

Hayward, A.D., Wilson, A.J., Pilkington, J.G., Pemberton, J.M. & Kruuk, L.E.B. (2009) Ageing in a variable habitat: environmental stress affects senescence in parasite resistance in St Kilda Soay sheep. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 3477–3485.

He, K., Liu, H.-Y., Ge, Y.-F., Wu, S.-Y. & Wan, Q.-H. (2017) Historical gene flow and profound spatial genetic structure among golden pheasant populations suggested by multi-locus analysis. *Molecular Phylogenetics and Evolution*, **110**, 93–103.

Hediger, R., Ansari, H.A. & Stranzinger, G.F. (1991) Chromosome banding and gene localizations support extensive conservation of chromosome structure between cattle and sheep. *Cytogenetic and Genome Research*, **57**, 127–134.

Hedrick, P.W. (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution*, **56**, 1902–1908.

Hedrick, P.W. (2004) Evolutionary genomics: foxy MHC selection story. *Heredity*, **93**, 237–238.

Hedrick, P.W. (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution*, **27**, 698–704.

Hedrick, P.W. & Black, F.L. (1997) HLA and Mate Selection: No Evidence in South Amerindians. *The American Journal of Human Genetics*, **61**, 505–511.

Hedrick, P.W. & Kim, T.J. (2000) Genetics of complex polymorphisms: Parasites and maintenance of the major histocompatibility variation. *Evolutionary Genetics: From Molecules to Morphology* (eds R.S. Singh & C.B. Krimbas), pp. 204–234. Cambridge University Press, Cambridge.

Hedrick, P.W., Parker, K.M. & Lee, R.N. (2001) Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology*, **10**, 1399–1412.

Hedrick, P.W., Thomson, G. & Klitz, W. (1987) Evolutionary genetics and HLA: another classic example PHILIP. *Biological Journal of the Linnean Society*, **31**, 311–331.

Herrmann-Hoesing, L.M., White, S.N., Kappmeyer, L.S., Herndon, D.R. & Knowles, D.P. (2008a) Genomic analysis of Ovis aries (Ovar) MHC class IIa loci. *Immunogenetics*, **60**, 167–176.

Herrmann-Hoesing, L.M., White, S.N., Mousel, M.R., Lewis, G.S. & Knowles, D.P. (2008b) Ovine progressive pneumonia provirus levels associate with breed and Ovar-DRB1. *Immunogenetics*, **60**, 749–758.

Hess, C.M. & Edwards, S. V. (2002) The evolution of the Major Histocompatibility complex in birds. *Bioscience*, **52**, 423–431.

Hess, C.M., Gasper, J., Hoekstra, H.E., Hill, C.E. & Edwards, S. V. (2000) MHC class II pseudogene and genomic signature of a 32-kb cosmid in the house finch (Carpodacus mexicanus). *Genome Research*, **10**, 613–23.

Hickford, J.G.H., Ridgway, H.J. & Escayg, A.P. (2000) Evolution of the ovine MHC DQA region. *Animal Genetics*, **31**, 200–205.

Hickford, J.G.H., Zhou, H. & Fang, Q. (2007) Haplotype analysis of the DQA genes in

sheep: Evidence supporting recombination between the loci. *Journal of Animal Science*, **85**, 577.

Hickford, J.G.H., Zhou, H., Slow, S. & Fang, Q. (2004) Diversity of the ovine DQA2 gene. *Journal of Animal Science*, **82**, 1553.

Hill, A.V.S. (1991) HLA associations with malaria in Africa: some implications for MHC evolution. *Molecular evolution of the major histocompatibility complex* (eds J. Klein & D. Klein), pp. 403–419. Springer, Berlin.

Hill, A.V.S. (1998) The immunogenetics of human infectious diseases. *Annu. Rev. Immunol*, **16**, 593–617.

Hill, R.E. & Hastie, N.D. (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature*, **326**, 96–99.

Hoelzel, A.R., Stephens, J.C. & O 'brien, S.J. (1999) Molecular Genetic Diversity and Evolution at the MHC DQB Locus in Four Species of Pinnipeds. *Mol. Biol. Evol*, **16**, 611–618.

Högstrand, K. & Böhme, J. (1999) Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. *Immunogenetics*, **49**, 446–55.

Holm, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Wright, M.W., Wain, H.M., Trowsdale, J., Ziegler, A. & Beck, S. (2004) Gene map of the extended human MHC. *Nature Reviews Genetics*, **5**, 889–899.

Huang, S.-W. & Yu, H.-T. (2003) Genetic Variation of Microsatellite Loci in the Major Histocompatibility Complex (MHC) Region in the Southeast Asian House Mouse (Mus musculus castaneus). *Genetica*, **119**, 201–218.

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L. a & Cowlishaw, G. (2010) A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology*, **10**, 96.

Huchard, E., Weill, M., Cowlishaw, G., Raymond, M. & Knapp, L.A. (2008) Polymorphism, haplotype composition, and selection in the Mhc-DRB of wild baboons. *Immunogenetics*, **60**, 585–598.

Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics Applications Note*, **17**, 754–755.

Hughes, A.L. & Hughes, M.K. (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics*, **42**, 233–243.

Hughes, A.L. & Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.

Hughes, A.L. & Nei, M. (1992) Models of host-parasite interaction and MHC polymorphism. *Genetics*, **132**, 863–864.

Hughes, A.L. & Yeager, M. (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual review of genetics*, **32**, 415–435.

Huisman, J. (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*.

References

Hunt, H.D., Pharr, G.T. & Bacon, L.D. (1994) Molecular Analysis Reveals Mhc Class-I Intra-Locus Recombination In The Chicken. *Immunogenetics*, **40**, 370–375.

Illius, A.W., Albon, S.D., Pemberton, J.M., Gordon, I.J. & Clutton-Brock, T.H. (1995) Selection for Foraging Efficiency During a Population Crash in Soay Sheep. *Journal of Animal Ecology*, **64**, 481–492.

Ilmonen, P., Penn, D.J., Damjanovich, K., Morrison, L., Ghotbi, L. & Potts, W.K. (2007) Major Histocompatibility Complex Heterozygosity Reduces Fitness in Experimentally Infected Mice. *Genetics*, **176**, 2501–2508.

Jaeger, C.P., Duvall, M.R., Swanson, B.J., Phillips, C.A., Dreslik, M.J., Baker, S.J. & King, R.B. (2016) Microsatellite and major histocompatibility complex variation in an endangered rattlesnake, the Eastern Massasauga (Sistrurus catenatus). *Ecology and Evolution*, **6**, 3991–4003.

Jeffery, K.J.M. & Bangham, C.R.M. (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.

Jeffreys, A., Kauppi, L. & Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature genetics*, **29**, 217–222.

Jensen, P.E. (2007) Recent advances in antigen processing and presentation. *Nature Immunology*, **8**, 1041–1048.

Jensen, L.F., Hansen, M.M., Mensberg, K.-L.D. & Loeschcke, V. (2008) Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (Salmo trutta, L.). *Heredity*, **100**, 79–91.

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., Muzny, D.M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W.C., Hourlier, T., Aken, B.L., Searle, S.M.J., Adelson, D.L., Bian, C., Cam, G.R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I.R., Fu, S., Guan, R., Highland, M.A., Holder, M.E., Huang, G., Ingham, A.B., Jhangiani, S.N., Kalra, D., Kovar, C.L., Lee, S.L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S.N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J.B., Kristensen, K., Gibbs, R.A., Flicek, P., Warkup, C.C., Jones, H.E., Oddy, V.H., Nicholas, F.W., McEwan, J.C., Kijas, J., Wang, J., Worley, K.C., Archibald, A.L., Cockett, N., Xu, X., Wang, W. & Dalrymple, B.P. (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science (New York, N.Y.)*, **344**, 1168–1173.

Johnston, S.E., Bérénos, C., Slate, J. & Pemberton, J.M. (2016) Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries). *Genetics*, **203**.

Johnston, S.E., Gratten, J., Berenos, C., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Slate, J. (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, **502**, 93–5.

Johnston, S.E., McEwan, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G., Pemberton, J.M. & Slate, J. (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, **20**, 2555–2566.

Jones, M.R., Cheviron, Z.A. & Carling, M.D. (2014) Variation in positively selected major

histocompatibility complex class I loci in rufous-collared sparrows (Zonotrichia capensis). *Immunogenetics*, **66**, 693–704.

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. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 2619–2625.

Jordan, W.C. & Bruford, M.W. (1998) New perspectives on mate choice and the MHC. *Heredity*, **81 ( Pt 3)**, 239–245.

Juola, F.A. & Dearborn, D.C. (2011) Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 153–162.

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T.B.., Sommerfeld, R.D., Wegner, K.M. & Milinski, M. (2009) Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**.

Kamath, P.L., Turner, W.C., Küsters, M. & Getz, W.M. (2014) Parasite-mediated selection drives an immunogenetic trade-off in plains zebras (Equus quagga). *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20140077.

Karl, J.A., Bohn, P.S., Wiseman, R.W., Nimityongskul, F.A., Lank, S.M., Starrett, G.J. & O'Connor, D.H. (2013) Major Histocompatibility Complex Class I Haplotype Diversity in Chinese Rhesus Macaques. *G3: Genes, Genomes, Genetics*, **3**, 1195–1201.

Karlsson, M., Schroeder, J., Nakagawa, S., Smith, H.G., Burke, T. & Westerdahl, H. (2015) House sparrow Passer domesticus survival is not associated with MHC-I diversity, but possibly with specific MHC-I alleles. *Journal of Avian Biology*, **46**, 167–174.

Kaufman, J., Milne, S., Göbel, T.W.F., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R. & Beck, S. (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, **401**, 923–925.

Keane, O.M., Dodds, K.G., Crawford, A.M. & McEwan, J.C. (2007) Transcriptional profiling of Ovis aries identifies Ovar-DQA1 allele frequency differences between nematode-resistant and susceptible selection lines. *Physiological Genomics*, **30**, 253–261.

Kekäläinen, J., Vallunen, J.A., Primmer, C.R., Rättyä, J. & Taskinen, J. (2009) Signals of major histocompatibility complex overdominance in a wild salmonid population. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 3133–3140.

Kelley, J., Walter, L. & Trowsdale, J. (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.

Kennedy, L.J., Modrell, A., Groves, P., Wei, Z., Single, R.M. & Happ, G.M. (2011) Genetic diversity of the major histocompatibility complex class II in Alaskan caribou herds. *International Journal of Immunogenetics*, **38**, 109–119.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. & Consortium, other members of the I.S.G. (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection (ed C Tyler-Smith). *PLoS Biology*, **10**, e1001258.

References

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kjøglum, S., Larsen, S., Bakke, H.G. & Grimholt, U. (2008) The Effect of Specific MHC Class I and Class II Combinations on Resistance to Furunculosis in Atlantic Salmon ( Salmo salar ). *Scandinavian Journal of Immunology*, **67**, 160–168.

Klein, J. (1986) *The Natural History of the Major Histocompatability Complex*. Wiley, New York, NY.

Klein, J. & O'hUigin, C. (1993) Composite origin of major histocompatibility complex genes. *Current opinion in genetics & development*, **3**, 923–30.

Kloch, A., Babik, W., Bajer, A., Siński, E. & Radwan, J. (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodes glareolus. *Molecular Ecology*, **19**, 255–265.

Kloch, A., Baran, K., Buczek, M., Konarzewski, M. & Radwan, J. (2013) MHC influences infection with parasites and winter survival in the root vole Microtus oeconomus. *Evolutionary Ecology*, **27**, 635–653.

Knafler, G.J., Clark, J.A., Boersma, P.D. & Bouzat, J.L. (2012) MHC diversity and mate choice in the magellanic penguin, Spheniscus magellanicus. *Journal of Heredity*, **103**, 759–768.

Kovacsovics-Bankowski, M. & Rock, K.L. (1995) A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science*, **267**, 243.

Kropshofer, H., Hämmerling, G.J. & Vogt, A.B. (1997) How HLA-DM edits the MHC class II peptide repertoire: survival of the fittest? *Immunology Today*, **18**, 77–82.

Kulski, J.K., Shiina, T., Anzai, T., Kohara, S. & Inoko, H. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews*, **190**, 95–122.

Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.

Kuroda, N., Figueroa, F., O'hUigin, C. & Klein, J. (2002) Evidence that the separation of Mhc class II from class I loci in the zebrafish, Danio rerio, occurred by translocation. *Immunogenetics*, **54**, 418–430.

Kurtz, J., Kalbe, M., Aeschlimann, P.B., Häberli, M.A., Wegner, K.M., Reusch, T.B.H. & Milinski, M. (2004) Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proceedings of the Royal Society of London B: Biological Sciences*, **271**, 197–204.

Kyle, C.J., Rico, Y., Castillo, S., Srithayakumar, V., Cullingham, C.I., White, B.N. & Pond, B.A. (2014) Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (Procyon lotor) populations exposed to raccoon rabies. *Molecular Ecology*, **23**, 2287–2298.

Lancaster, A.K., Single, R.M., Solberg, O.D., Nelson, M.P. & Thomson, G. (2007) PyPop update - A software pipeline for large-scale multilocus population genomics. *Tissue Antigens*, **69**, 192–197.

Landry, C. & Bernatchez, L. (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and

microsatellite loci in Atlantic salmon (Salmo salar). *Molecular Ecology*, **10**, 2525–2539.

Langefors, Å.H. (2005) Adaptive and neutral genetic variation and colonization history of Atlantic salmon, Salmo salar. *Environmental Biology of Fishes*, **74**, 297–308.

Langefors, Å.H., Lohm, J., Grahn, M., Andersen, Ø. & Schantz, T. von. (2001a) Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 479–485.

Langefors, A., Lohm, J. & von Schantz, T. (2001b) Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (Salmo salar). *Immunogenetics*, **53**, 329–36.

Larruskain, A., Minguijón, E., García-Etxebarria, K., Moreno, B., Arostegui, I., Juste, R.A. & Jugo, B.M. (2010) MHC class II DRB1 gene polymorphism in the pathogenesis of Maedi–Visna and pulmonary adenocarcinoma viral diseases in sheep. *Immunogenetics*, **62**, 75–83.

Lawlor, D.A., Zemmour, J., Ennis, P.D. & Parham, P. (1990) Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. , **8**, 23–63.

Lawrence, C.E., Kennedy, M.W. & Garside, P. (2013) Gut Immunopathology in Helminth Infections - Paradigm Lost? *Parasitic Nematodes: Molecular Biology, Biochemistry and Immunology*, 2nd ed (eds M.W. Kennedy & W. Harnett), pp. 374–398. CABI, Wallingford.

Lee, C.Y., Munyard, K.A., Gregg, K., Wetherall, J.D., Stear, M.J. & Groth, D.M. (2011) The influence of MHC and immunoglobulins A and E on host resistance to gastrointestinal nematodes in sheep. *Journal of Parasitology Research*, **2011**, 101848.

Lee, C.Y., Qin, J., Munyard, K.A., Siva Subramaniam, N., Wetherall, J.D., Stear, M.J. & Groth, D.M. (2012) Conserved haplotype blocks within the sheep MHC and low SNP heterozygosity in the Class IIa subregion. *Animal Genetics*, **43**, 429–437.

Lei, W., Fang, W., Lin, Q., Zhou, X. & Chen, X. (2015) Characterization of a non-classical MHC class II gene in the vulnerable Chinese egret (Egretta eulophotes). *Immunogenetics*, **67**, 463–472.

Lenz, T.L., Eizaguirre, C., Becker, S. & Reusch, T.B. (2009a) RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback Gasterosteus aculeatus. *BMC Evolutionary Biology*, **9**, 57.

Lenz, T.L., Eizaguirre, C., Scharsack, J.P., Kalbe, M. & Milinski, M. (2009b) Disentangling the role of MHC-dependent “good genes” and “compatible genes” in mate-choice decisions of three-spined sticklebacks Gasterosteus aculeatus under semi-natural conditions. *Journal of Fish Biology*, **75**, 2122–2142.

Lenz, T.L., Spirin, V., Jordan, D.M. & Sunyaev, S.R. (2016) Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. *Molecular biology and evolution*, **33**, 2555–2564.

Lenz, T.L., Wells, K., Pfeiffer, M. & Sommer, S. (2009c) Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the Long-tailed giant rat (Leopoldamys sabanus). *BMC Evolutionary Biology*, **9**, 269.

Lewontin, R.C., Ginzburgh, L.R. & Tuljapurkar, S.D. (1978) Heterosis as an explanation for large amounts of genic polymorphism. *Genetics*, **88**, 149–170.

References

Li, L., Wang, B.B., Ge, Y.F. & Wan, Q.H. (2014) Major histocompatibility complex class II polymorphisms in forest musk deer (Moschus berezovskii) and their probable association with purulent disease. *International Journal of Immunogenetics*, 401–412.

Lighten, J., van Oosterhout, C., Paterson, I.G., Mcmullan, M. & Bentzen, P. (2014) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (Poecilia reticulata). *Molecular Ecology Resources*, **14**, 753–767.

Lillie, M., Grueber, C.E., Sutton, J.T., Howitt, R., Bishop, P.J., Gleeson, D. & Belov, K. (2015) Selection on MHC class II supertypes in the New Zealand endemic Hochstetter's frog. *BMC Evolutionary Biology*, **15**, 1–11.

Liu, H., Liu, K., Wang, J. & Ma, R.Z. (2006) A BAC clone-based physical map of ovine major histocompatibility complex. *Genomics*, **88**, 88–95.

Liu, H.-Y., Xue, F., Wan, Q.-H. & Ge, Y.F. (2013) MHC class II β genes in the endangered Hainan Eld's Deer (Cervus eldi hainanus). *Journal of Heredity*, **104**, 874–880.

Liu, K., Zhang, P., Gao, J., Liu, H., Li, G., Qiu, Z., Zhang, Y., Ren, J., Tan, P. & Ma, R.Z. (2011) Closing a gap in the physical map of the ovine major histocompatibility complex. *Animal Genetics*, **42**, 204–207.

Lizée, G., Basha, G., Tiong, J., Julien, J.-P., Tian, M., Biron, K.E. & Jefferies, W.A. (2003) Control of dendritic cell cross-presentation by the major histocompatibility complex class I cytoplasmic domain. *Nature Immunology*, **4**, 1065–1073.

Llaurens, V., Billiard, S., Leducq, J.B., Castric, V., Klein, E.K. & Vekemans, X. (2008) Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in Arabidopsis halleri? *Evolution*, **62**, 2545–2557.

Lochmiller, R.L. & Deerenberg, C. (2000) Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, **88**, 87–98.

Lohm, J., Grahn, M., Langefors, Å.H., Andersen, Ø., Storset, A. & Schantz, T. von. (2002) Experimental evidence for major histocompatibility complex–allele–specific resistance to a bacterial infection. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 2029–2033.

Loiseau, C., Richard, M., Garnier, S., Chastel, O., Julliard, R., Zoorob, R. & Sorci, G. (2009) Diversifying selection on MHC class I in the house sparrow ( Passer domesticus ). *Molecular Ecology*, **18**, 1331–1340.

Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R. & Sorci, G. (2008) Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. *Ecology Letters*, **11**, 258–265.

Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. & Sorci, G. (2011) Plasmodium relictum infection and MHC diversity in the house sparrow (Passer domesticus). *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 1264–1272.

Løvlie, H., Gillingham, M.A.F., Worley, K., Pizzari, T. & Richardson, D.S. (2013) Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20131296.

Lukasch, B., Westerdahl, H., Strandh, M., Knauer, F., Winkler, H., Moodley, Y. & Hoi, H.

(2017) Major histocompatibility complex genes partly explain early survival in house sparrows. *Scientific Reports*, **7**, 6571.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Røder, G., Justesen, S., Buus, S. & Brunak, S. (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*, **55**, 797–810.

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 Research*, **11**, 1275–89.

Madsen, T. & Ujvari, B. (2006) MHC class I variation associates with parasite resistance and longevity in tropical pythons. *Journal of Evolutionary Biology*, **19**, 1973–1978.

MAFF. (1986) *Manual of Veterinary Parasitological Laboratory Techniques*. London.

Mahdy, A., Makinen, A., Chowdhary, B.P., Andersson, L. & Gustavsson, I. (1989) Chromosomal localization of the ovine major histocompatibility complex (OLA) by in situ hybridization. *Hereditas*, **111**, 87–90.

Maizels, R.M. & Yazdanbakhsh, M. (2003) Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nature Reviews. Immunology*, **3**, 733–744.

Mann, A., Abraham, L., Cameron, P., Robinson, W., Giphart, M.J. & Dawkins, R. (1993) The caprine MHC contains DYA genes. *Immunogenetics*, **37**, 292–295.

Marcos, C.Y., Fernández-Vina, M.A., Lázaro, A.M., Nulf, C.J., Raimondi, E.H. & Stastny, P. (1997) Novel HLA-B35 subtypes. *Human Immunology*, **53**, 148–155.

Markow, T., Hedrick, P.W., Zuerlein, K., Danilovs, J., Martin, J., Vyvial, T. & Armstrong, C. (1993) HLA polymorphism in the Havasupai: evidence for balancing selection. *American journal of human genetics*, **53**, 943–52.

Marsh, S.G.E., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Fernández-Viña, M., Geraghty, D.E., Holdsworth, R., Hurley, C.K., Lau, M., Lee, K.W., Mach, B., Maiers, M., Mayr, W.R., Müller, C.R., Parham, P., Petersdorf, E.W., Sasazuki, T., Strominger, J.L., Svejgaard, A., Terasaki, P.I., Tiercy, J.M. & Trowsdale, J. (2010) Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, **75**, 291–455.

Martin, A., Dunnington, E.A., Briles, W.E., Briles, R.W. & Siegel, P.B. (1989) Marek's disease and major histocompatibility complex haplotypes in chickens selected for high or low antibody response. *Animal Genetics*, **20**, 407–414.

Martinez-Valladares, M., Vara-Del Rio, M.P., Cruz-Rojo, M.A. & Rojo-Vazquez, F.A. (2005) Genetic resistance to Teladorsagia circumcincta: IgA and parameters at slaughter in Churra sheep. *Parasite Immunology*, **27**, 213–218.

Maruyama, T. & Nei, M. (1981) Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics*, **98**, 441–459.

Massingham, T. & Goldman, N. (2005) Detecting Amino Acid Sites Under Positive Selection and Purifying Selection. *Genetics*, **169**, 1753–1762.

McAdam, S.N., Boyson, J.E., Liu, X., Garber, T.L., Hughes, A.L., Bontrop, R.E. & Watkins,

D.I. (1994) A uniquely high level of recombination at the HLA-B locus. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 5893–7.

McCairns, R.J.S., Bourget, S. & Bernatchez, L. (2011) Putative causes and consequences of MHC variation within and between locally adapted stickleback demes. *Molecular Ecology*, **20**, 486–502.

McClelland, E.E., Penn, D.J. & Potts, W.K. (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infection and immunity*, **71**, 2079–86.

Meyer-Lucht, Y. & Sommer, S. (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (Apodemus flavicollis). *Molecular Ecology*, **14**, 2233–2243.

Meyer, D. & Thomson, G. (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1–26.

Milinski, M. (2006) The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 159–186.

Milinski, M., Croy, I., Hummel, T. & Boehm, T. (2013) Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20122889.

Miller, H.C., Allendorf, F. & Daugherty, C.H. (2010) Genetic diversity and differentiation at MHC genes in island populations of tuatara (Sphenodon spp.). *Molecular Ecology*, **19**, 3894–3908.

Miller, K.M., Kaukinen, K.H., Beacham, T.D. & Withler, R.E. (2001) Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica*, **111**, 237–257.

Miller, H.C. & Lambert, D.M. (2004a) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics*, **56**, 178–191.

Miller, H.C. & Lambert, D.M. (2004b) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology*, **13**, 3709–3721.

Miller, H.C., Moore, J.A., Nelson, N.J. & Daugherty, C.H. (2009) Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1695–1704.

Miller, K.M. & Withler, R.E. (1997) Mhc diversity in Pacific salmon: population structure and trans-species allelism. *Hereditas*, **127**, 83–95.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F. & Wright, F. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*, **25**, 126–127.

Milner, J.M., Albon, S.D., Illius, A.W., Pemberton, J.M. & Clutton-Brock, T.H. (1999) Repeated selection of morphometric traits in the Soay sheep on St Kilda. *Journal of Animal Ecology*, **68**, 472–488.

Miltiadou, D., Ballingall, K.T., Ellis, S.A., Russell, G.C. & McKeever, D.J. (2005) Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. *Immunogenetics*, **57**, 499–509.

Miyasaka, T., Takeshima, S.-N., Sentsui, H. & Aida, Y. (2012) Identification and diversity

of bovine major histocompatibility complex class II haplotypes in Japanese Black and Holstein cattle in Japan. *Journal of Dairy Science*, **95**, 420–431.

Muirhead, C.A. (2001) Consequences of population structure on genes under balancing selection. *Evolution*, **55**, 1532–1541.

Murphy, L., Eckersall, P.D., Bishop, S.C., Pettit, J.J., Huntley, J.F., Burchmore, R. & Stear, M.J. (2010) Genetic variation among lambs in peripheral IgE activity against the larval stages of Teladorsagia circumcincta. *Parasitology*, **137**, 1249–1260.

Murphy, K., Travers, P., Walport, M. & Janeway, C. (2012) *Janeway's Immunobiology*, 8th ed. Garland Science, New York.

Musolf, K., Meyer-Lucht, Y. & Sommer, S. (2004) Evolution of MHC-DRB class II polymorphism in the genus Apodemus and a comparison of DRB sequences within the family Muridae (Mammalia: Rodentia). *Immunogenetics*, **56**, 420–426.

Naugler, C. & Liwski, R. (2008) An evolutionary approach to major histocompatibility diversity based on allele supertypes. *Medical Hypotheses*, **70**, 933–937.

Navarro, A. & Barton, N.H. (2002) The effects of multilocus balancing selection on neutral variability. *Genetics*, **161**, 849–863.

Neefjes, J., Jongsma, M.L.M., Paul, P. & Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, **11**, 823–836.

Nei, M., Gu, X. & Sitnikova, T. (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS*, **94**, 7799–7806.

Nei, M. & Hughes, A.L. (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. *Proceedings of the 11th histocompatibility workshop and conference*, pp. 27–38. Oxford University Press, Oxford.

Nei, M. & Rooney, A.P. (2005) Concerted and birth-and-death evolution of multigene families. *Annual review of genetics*, **39**, 121–152.

De Nettancourt, D. (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants*. Springer, Berlin.

Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641–647.

Nielsen, R. & Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**, 929–36.

Nieuwhof, G.J. & Bishop, S.C. (2005) Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. *Animal Science*, **81**.

Niskanen, A.K., Kennedy, L.J., Lohi, H., Aspi, J. & Pyhäjärvi, T. (2016) No evidence of prenatal diversifying selection at locus or supertype levels in the dog MHC class II loci. *Canine Genetics and Epidemiology*, **3**, 9.

Niskanen, A.K., Kennedy, L.J., Ruokonen, M., Kojola, I., Lohi, H., Isomursu, M., Jansson, E., Pyhäjärvi, T. & Aspi, J. (2014) Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population. *Molecular Ecology*, **23**, 875–889.

Norbury, C.C., Hewlett, L.J., Prescott, A.R., Shastri, N. & Watts, C. (1995) Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow

macrophages. *Immunity*, **3**, 783–791.

Nowak, M.A., Tarczy-Hornoch, K. & Austyn, J.M. (1992) The optimal number of major histocompatibility complex molecules in an individual. *PNAS*, **89**, 10896–10899.

Nussey, D.H., Watt, K.A., Clark, A., Pilkington, J.G., Pemberton, J.M., Graham, A.L. & Mcneilly, T.N. (2014) Multivariate immune defences and fitness in the wild: complex antibodies, health and survival. *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20132931.

O'Farrell, B., Benzie, J.A.H., McGinnity, P., Carlsson, J., Eyto, E. de, Dillane, E., Graham, C., Coughlan, J. & Cross, T. (2012) MHC-mediated spatial distribution in brown trout (Salmo trutta) fry. *Heredity*, **108**, 403–409.

Ober, C., Weitkamp, L.R., Cox, N., Dytch, H., Kostyu, D. & Elias, S. (1997) HLA and mate choice in humans. *American journal of human genetics*, **61**, 497–504.

Ohta, T. (1999) Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunological Reviews*, **167**, 319–25.

Ohta, Y., Okamura, K., McKinney, E.C., Bartl, S., Hashimoto, K. & Flajnik, M.F. (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 4712–7.

Oliver, M.K., Lambin, X., Cornulier, T. & Piertney, S.B. (2009a) Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (Arvicola terrestris) metapopulations. *Molecular Ecology*, **18**, 80–92.

Oliver, M.K. & Piertney, S.B. (2006) Isolation and characterization of a MHC class II DRB locus in the European water vole (Arvicola terrestris). *Immunogenetics*, **58**, 390–395.

Oliver, M.K. & Piertney, S.B. (2012) Selection maintains mhc diversity through a natural population bottleneck. *Molecular Biology and Evolution*, **29**, 1713–1720.

Oliver, M.K., Telfer, S. & Piertney, S.B. (2009b) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 1119–1128.

Olsson, M., Madsen, T., Wapstra, E., Silverin, B., Ujvari, B. & Wittzell, H. (2005) MHC, health, color, and reproductive success in sand lizards. *Behavioral Ecology and Sociobiology*, **58**, 289–294.

van Oorschot, R.A.H., Maddox, J.F., Adams, L.A. & Fabb, S.A. (1994) Characterization and evolution of ovine MHC class II DQB sequence polymorphism. *Animal Genetics*, **25**, 417–424.

van Oosterhout, C. (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 657–665.

van Oosterhout, C., Joyce, D.A., Cummings, S.M., Blais, J., Barson, N.J., Ramnarine, I.W., Mohammed, R.S., Persad, N. & Cable, J. (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (Poecilia reticulata). *Evolution*, **60**, 2562–2574.

Oppelt, C., Starkloff, A., Rausch, P., Von Holst, D. & Rödel, H.G. (2010) Major

histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. *Molecular Ecology*, **19**, 4155–4167.

Osborne, A.J., Pearson, J., Negro, S.S., Chilvers, B.L., Kennedy, M. a. & Gemmell, N.J. (2015) Heterozygote advantage at MHC DRB may influence response to infectious disease epizootics. *Molecular Ecology*, **24**, 1419–1432.

Osborne, M.J., Pilger, T.J., Lusk, J.D. & Turner, T.F. (2017) Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class IIβ in the endangered Rio Grande silvery minnow. *Molecular Ecology*, **26**, 471–489.

Ota, T. & Nei, M. (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular biology and evolution*, **11**, 469–482.

Ou, D., Mitchell, L.A. & Tingle, A.J. (1998) A new categorization of HLA DR alleles on a functional basis. *Human Immunology*, **59**, 665–676.

Outteridge, P.M., Andersson, L., Douch, P.G., Green, R.S., Gwakisa, P.S., Hohenhaus, M.A. & Mikko, S. (1996) The PCR typing of MHC-DRB genes in the sheep using primers for an intronic microsatellite: application to nematode parasite resistance. *Immunology and cell biology*, **74**, 330–336.

Owen, J.P., Delany, M.E. & Mullens, B. a. (2008) MHC haplotype involvement in avian resistance to an ectoparasite. *Immunogenetics*, **60**, 621–631.

Parham, P. & Ohta, T. (1996) Population Biology of Antigen Presentation by MHC Class I Molecules. *Science*, **272**, 67–74.

Paterson, S. (1996) *Major Histocompatibility Complex Variation in Soay Sheep*. University of Cambridge.

Paterson, S. (1998) Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *Journal of Heredity*, **89**, 289–294.

Paterson, S. & Pemberton, J.M. (1997) No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proceedings of the Royal Society of London B: Biological Sciences*, **264**, 1813–1819.

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. *PNAS*, **95**, 3714–3719.

Pavey, S., Sevellec, M., Adam, W., Normandeau, E., Lamaze, F.C., Gagnaire, P.A., Filteau, M., Hebert, F.O., Maaroufi, H. & Bernatchez, L. (2013) Nonparallelism in MHCIIB diversity accompanies nonparallelism in pathogen infection of lake whitefish (Coregonus clupeaformis) species pairs as revealed by next-generation sequencing. *Molecular Ecology*, **22**, 3833–3849.

Pearson, S.K., Bull, C.M. & Gardner, M.G. (2017) Egernia stokesii (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. *Immunogenetics*, **69**, 49–61.

Pečnerová, P., Díez-del-Molino, D., Vartanyan, S. & Dalén, L. (2016) Changes in variation at the MHC class II DQA locus during the final demise of the woolly mammoth. *Scientific Reports*, 25274.

Penn, D.J. (2002) The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.

References


Penn, D.J., Damjanovich, K. & Potts, W.K. (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *PNAS*, **99**, 11260–11264.

Penn, D.J. & Potts, W.K. (1999) The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist*, **153**, 145–164.

Piertney, S.B. & Oliver, M.K. (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Piontkivska, H. & Nei, M. (2003) Birth-and-Death Evolution in Primate MHC Class I Genes: Divergence Time Estimates. *Molecular Biology and Evolution*, **20**, 601–609.

van der Poel, J.J., Groenen, M.A.M., Dijkhof, R.J.M., Ruyter, D. & Giphart, M.J. (1990) The nucleotide sequence of the bovine MHC class II alpha genes:DRA, DQA, andDYA. *Immunogenetics*, **31**, 29–36.

Potts, W.K. & Wakeland, E.K. (1990) Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, **5**, 181–187.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, **81**, 559–575.

Qin, J., Mamotte, C., Cockett, N.E., Wetherall, J.D. & Groth, D.M. (2008) A map of the class III region of the sheep major histocompatibilty complex. *BMC Genomics*, **9**, 409.

R Core Team. (2017) R: A language and environment for statistical computing.

Radwan, J., Kuduk, K., Levy, E., Lebas, N. & Babik, W. (2014) Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Molecular Ecology*, **23**, 5966–5978.

Radwan, J., Tkacz, A. & Kloch, A. (2008) MHC and Preferences for Male Odour in the Bank Vole. *Ethology*, **114**, 827–833.

Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L. & Babik, W. (2012) MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology*, **21**, 2469–2479.

Raymond, M. & Rousset, F. (1995) An Exact Test for Population Differentiation. *Evolution*, **49**, 1280–1283.

Reche, P.A. & Reinherz, E.L. (2003) Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology*, **331**, 623–641.

Reddacliff, L., Beh, K., McGregor, H. & Whittington, R. (2005) A preliminary study of possible genetic influences on the susceptibility of sheep to Johne's disease. *Australian Veterinary Journal*, **83**, 435–441.

Reed, D.H. & Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Reichard, M., Spence, R., Bryjová, A., Bryja, J. & Smith, C. (2012) Female Rose Bitterling Prefer MHC-Dissimilar Males: Experimental Evidence. *PLoS ONE*, **7**, e40780.

Reusch, T.B.H., Häberli, M.A., Aeschlimann, P.B. & Milinski, M. (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, **414**, 300–302.

Reusch, T.B.H. & Langefors, Å. (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback Gasterosteus aculeatus. *Journal of Molecular Evolution*, **61**, 531–541.

Rhoads, A. & Au, K.F. (2015) PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, **13**, 278–289.

Richardson, D.S. & Westerdahl, H. (2003) MHC diversity in two Acrocephalus species: The outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology*, **12**, 3523–3529.

Riggio, V., Pong-Wong, R., Sallé, G., Usai, M.G., Casu, S., Moreno, C.R., Matika, O. & Bishop, S.C. (2014) A joint analysis to identify loci underlying variation in nematode resistance in three European sheep populations. *Journal of Animal Breeding and Genetics*, 1–11.

Robinson, M.R., Wilson, A.J., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Kruuk, L.E.B. (2009) The impact of environmental heterogeneity on genetic architecture in a wild population of soay sheep. *Genetics*, **181**, 1639–1648.

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ruan, R., Ruan, J., Wan, X.-L., Zheng, Y., Chen, M.-M., Zheng, J.-S. & Wang, D. (2016) Organization and characteristics of the major histocompatibility complex class II region in the Yangtze finless porpoise (Neophocaena asiaeorientalis asiaeorientalis). *Scientific Reports*, **6**, 22471.

Ruiz-López, M.J., Monello, R.J., Schuttler, S.G., Lance, S.L., Gompper, M.E. & Eggert, L.S. (2014) Major Histocompatibility Complex, demographic, and environmental predictors of antibody presence in a free-ranging mammal. *Infection, Genetics and Evolution*, **28**, 317–327.

Salmier, A., de Thoisy, B., Crouau-Roy, B., Lacoste, V. & Lavergne, A. (2016) Spatial pattern of genetic diversity and selection in the MHC class II DRB of three Neotropical bat species. *BMC Evolutionary Biology*, **16**, 229.

Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M. & Wold, S. (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry*, **41**, 2481–2491.

Sato, A., Figueroa, F., Murray, B.W., Málaga-Trillo, E., Zaleska-Rutczynska, Z., Sültmann 7, H., Toyosawa, S., Wedekind, C., Steck, N. & Klein, J. (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics*, **51**, 108–116.

Savage, A.E. & Zamudio, K.R. (2011) MHC genotypes associate with resistance to a frog-killing fungus. *PNAS*, **108**, 16705–10.

Sayers, G., Good, B., Hanrahan, J.P., Ryan, M., Angles, J.M. & Sweeney, T. (2004) Major Histocompatibility Complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Heredity*, **92**, 102–109.

Sayers, G., Good, B., Hanrahan, J.P., Ryan, M., Angles, J.M. & Sweeney, T. (2005) Major Histocompatibility Complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology*, **131**, 403–409.

Schad, J., Ganzhorn, J.U. & Sommer, S. (2005) Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, Microcebus murinus.

*Evolution*, **59**, 439–450.

Schaschl, H., Suchentrunk, F., Hammer, S. & Goodman, S.J. (2005) Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (Rupicapra spp.). *Immunogenetics*, **57**, 108–115.

Schierup, M.H. & Vekemans, X. (2008) Genomic consequences of selection on self-incompatibility genes. *Current Opinion in Plant Biology*, **11**, 116–122.

Schierup, M.H., Vekemans, X. & Charlesworth, D. (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical research*, **76**, 51–62.

Schmid-Hempel, P. (2003) Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 357–366.

Schmid-Hempel, P. (2011) *Evolutionary Parasitology: The Integrated Study of Infections, Immunology, Ecology, and Genetics*. Oxford University Press, Oxford.

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

Schwensow, N., Dausmann, K., Eberle, M., Fietz, J. & Sommer, S. (2010) Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. *Infection, Genetics and Evolution*, **10**, 662–668.

Schwensow, N., Fietz, J., Dausmann, K.H. & Sommer, S. (2007) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity*, **99**, 265–277.

Scott, P.C., Choi, C.-L. & Brandon, M.R. (1987) Genetic organization of the ovine MHC class II region. *Immunogenetics*, **25**, 116–122.

Scott, P.C., Maddox, J.F., Gogolin-Ewens, K. & Brandon, M.R. (1991) The nucleotide sequence and evolution of ovine MHC class II B genes: DQB and DRB. *Immunogenetics*, **35**, 217.

Seifertová, M., Jarkovský, J. & Šimková, A. (2016) Does the parasite-mediated selection drive the MHC class IIB diversity in wild populations of European chub (Squalius cephalus)? *Parasitology Research*, **115**, 1401–1415.

Sepil, I., Lachish, S., Hinks, A.E. & Sheldon, B.C. (2013a) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20130134.

Sepil, I., Lachish, S. & Sheldon, B.C. (2013b) Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular Ecology*, **22**, 384–396.

Sepil, I., Moghadam, H.K., Huchard, E. & Sheldon, B.C. (2012) Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology*, **12**, 68.

Sette, A. & Sidney, J. (1998) HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. *Current Opinion in Immunology*, **10**, 478–482.

Sette, A. & Sidney, J. (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, **50**, 201–212.

She, J.X., Boehme, S.A., Wang, T.W., Bonhomme, F. & Wakeland, E.K. (1991) Amplification of Major Histocompatibility Complex Class II Gene Diversity by Intraexonic Recombination Source : Proceedings of the National Academy of Sciences of the United States of America , Published by : National Academy of Sciences Stable URL : http. *PNAS*, **88**, 453–457.

Shen, H., Han, G., Jia, B., Jiang, S. & Du, Y. (2014) MHC-DRB1/DQB1 gene polymorphism and its association with resistance/susceptibility to cystic echinococcosis in Chinese Merino sheep. *Journal of Parasitology Research*, **2014**, 272601.

Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M. a, Hedrick, R.P., Nehring, R.B., Stet, R.J., Secombes, C. & Parham, P. (2001) Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *Journal of immunology (Baltimore, Md. : 1950)*, **166**, 3297–3308.

Sidney, J., Grey, H.M., Kubo, R.T. & Sette, A. (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. *Immunology Today*, **17**, 261–266.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., Sö Ding, J., Thompson, J.D. & Higgins, D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539.

Silva, N.F. & Goring, D.R. (2001) Cellular and Molecular Life Sciences Mechanisms of self-incompatibility in flowering plants. *Cellular and Molecular Life Sciences2*, **58**, 1988–2007.

Sin, Y.W., Annavi, G., Dugdale, H.L., Newman, C., Burke, T. & MacDonald, D.W. (2014) Pathogen burden, co-infection and major histocompatibility complex variability in the European badger ( Meles meles ). *Molecular Ecology*, **23**, 5072–5088.

Siva Subramaniam, N., Morgan, E., Wetherall, J.D., Stear, M.J. & Groth, D. (2015) A comprehensive mapping of the structure and gene organisation in the sheep MHC class I region. *BMC Genomics*, **16**, 810.

Slade, R.W. & McCallum, H.I. (1992) Frequency-Dependent Selection at MHC Loci. *Genetics*, **132**, 861–862.

Slatkin, M. (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical research*, **64**, 71–74.

Slatkin, M. (1996) A correction to the exact test based on the Ewens sampling distribution. *Genetics Research*, **68**, 259–260.

Smith, S., De Bellocq, J.G., Suchentrunk, F. & Schaschl, H. (2011) Evolutionary genetics of MHC class II beta genes in the brown hare, Lepus europaeus. *Immunogenetics*, **63**, 743–751.

Solberg, O.D., Mack, S.J., Lancaster, A.K., Single, R.M., Tsai, Y., Sanchez-Mazas, A. & Thomson, G. (2008) Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Human Immunology*, **69**, 443–64.

Sommer, S. (2003) Effects of habitat fragmentation and changes of dispersal behaviour

after a recent population decline on the genetic variability of noncoding and coding DNA of a monogamous Malagasy rodent. *Molecular Ecology*, **12**, 2845–2851.

Sommer, S. (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, **2**, 16.

Spurgin, L.G., Van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C. & Richardson, D.S. (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20**, 5213–5225.

Spurgin, L.G. & Richardson, D.S. (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 979–988.

Stear, M.J., Abuagob, O., Benothman, M., Bishop, S.C., Innocent, G.T.T., Kerr, A. & Mitchell, S. (2006) Variation among faecal egg counts following natural nematode infection in Scottish Blackface lambs. *Parasitology*, **132**, 275–280.

Stear, M.J., Bairden, K., Bishop, S.C., Buitkamp, J., Epplen, J.T., Gostomski, D., McKellar, Q.A., Schwaiger, F.W. & Wallace, D.S. (1996) An ovine lymphocyte antigen is associated with reduced faecal egg counts in four-month-old lambs following natural, predominantly Ostertagia circumcincta infection. *International Journal for Parasitology*, **26**, 423–428.

Stear, M.J., Bairden, K., Innocent, G.T.T., Mitchell, S., Strain, S. & Bishop, S.C. (2004) The relationship between IgA activity against 4th-stage larvae and density-dependent effects on the number of 4th-stage larvae of Teladorsagia circumcincta in naturally infected sheep. *Parasitology*, **129**, 363–369.

Stear, M.J., Belch, A., Donskow-Schmelter, K., Fitton, L.A.A., Innocent, G.T.T., Ishikane, C., Mateus, A., Murphy, L., Rennie, K., Smith, A. & Sayers, G. (2007) Detection of genes with moderate effects on disease resistance using ovine mhc and resistance to nematodes as an example. *Veterinary Immunology and Immunopathology*, **120**, 3–9.

Stear, M.J., Innocent, G.T.T. & Buitkamp, J. (2005) The evolution and maintenance of polymorphism in the major histocompatibility complex. *Veterinary Immunology and Immunopathology*, **108**, 53–57.

Stear, M.J., Strain, S. & Bishop, S.C. (1999a) Mechanisms underlying resistance to nematode infection. *International Journal for Parasitology*, **29**, 51–56.

Stear, M.J., Strain, S. & Bishop, S.C. (1999b) How lambs control infection with Ostertagia circumcincta. *Veterinary Immunology and Immunopathology*, **72**, 213–218.

Stefan, T. (2016) *Mathematical Modelling of the Selective Forces Maintaining Diversity at the Major Histocompatibility Complex*. University of Glasgow.

Sterck, E.H.M., Bontrop, R.E., de Groot, N., de Vos-Rouweler, A.J.M. & Doxiadis, G.G.M. (2017) No postcopulatory selection against MHC-homozygous offspring: Evidence from a pedigreed captive rhesus macaque colony. *Molecular Ecology*, **26**, 3785–3793.

Stoffels, R.J. & Spencer, H.G. (2008) An asymmetric model of heterozygote advantage at major histocompatibility complex genes: degenerate pathogen recognition and intersection advantage. *Genetics*, **178**, 1473–1489.

Stone, R.T. & Muggli-Cockett, N.E. (1990) Partial nucleotide sequence of a novel bovine major histocompatibility complex class II β-chain gene, BoLA-DIB. *Animal Genetics*, **21**, 353–360.

Strain, S. a J., Bishop, S.C., Henderson, N.G., Kerr, A., McKellar, Q.A., Mitchell, S. & Stear, M.J. (2002) The genetic control of IgA activity against Teladorsagia circumcincta and its association with parasite resistance in naturally infected sheep. *Parasitology*, **124**, 545–552.

Strandh, M., Westerdahl, H., Pontarp, M., Canbäck, B., Dubois, M.-P., Miquel, C., Taberlet, P. & Bonadonna, F. (2012) Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4457–4463.

Subramaniam, R., White, S.N., Herrmann-Hoesing, L.M. & Srikumaran, S. (2012) MHC class II DR allelic diversity in bighorn sheep. *Gene*, **506**, 217–222.

Sutherland, I. & Scott, I. (2009) *Gastrointestinal Nematodes of Sheep and Cattle: Biology and Control*. John Wiley & Sons, Chichester.

Suzuki, Y. (2004) New Methods for Detecting Positive Selection at Single Amino Acid Sites. *Journal of Molecular Evolution*, **59**, 11–19.

Swiderek, W.P., Charon, K.M., Winnicka, a & Gruszczynska, J. (2006) Relationship between blood lymphocyte phenotype, DRB1 (MHC class II) gene polymorphism and somatic cell count in ewe milk. *Bulletin of the Veterinary Institute in Pulawy*, **50**, 73–77.

Takahata, N. & Nei, M. (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967–978.

Teacher, A.G.F., Garner, T.W.J. & Nichols, R.A. (2009) Evidence for directional selection at a novel major histocompatibility class I marker in wild common frogs (Rana temporaria) exposed to a viral pathogen (Ranavirus). *PLoS ONE*, **4**, 1–6.

Thursz, M.R., Thomas, H.C., Greenwood, B.M. & Hill, A.V.S. (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics*, **17**, 11–12.

Ting, J.P.-Y. & Trowsdale, J. (2002) Genetic Control of MHC Class II Expression. *Cell*, **109**, S21–S33.

Tison, J.L., Blennow, V., Palkopoulou, E., Gustafsson, P., Roos, A. & Dalén, L. (2015) Population structure and recent temporal changes in genetic variation in Eurasian otters from Sweden. *Conservation Genetics*, **16**, 371–384.

Tobler, M., Plath, M., Riesch, R., Schlupp, I., Grasse, A., Munimanda, G.K., Setzer, C., Penn, D.J. & Moodley, Y. (2014) Selection from parasites favours immunogenetic diversity but not divergence among locally adapted host populations. *Journal of Evolutionary Biology*, **27**, 960–974.

Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Chaval, Y., Berthier, K., Ribas Salvador, A., Voutilainen, L., Laakkonen, J., Henttonen, H., Cosson, J.F. & Charbonnel, N. (2008) Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: Insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology*, **21**, 1307–1320.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser,

R., Fugate, M., Theiler, J., Hsu, Y.S., Kunstman, K., Wu, S., Phair, J., Erlich, H. & Wolinsky, S. (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, **9**, 928–935.

Tracy, K.E., Kiemnec-Tyburczy, K.M., DeWoody, J.A., Parra-Olea, G. & Zamudio, K.R. (2015) Positive selection drives the evolution of a major histocompatibility complex gene in an endangered Mexican salamander species complex. *Immunogenetics*, **67**, 323–335.

Traherne, J.A., Horton, R., Roberts, A.N., Miretti, M.M., Hurles, M.E., Stewart, C.A., Ashurst, J.L., Atrazhev, A.M., Coggill, P., Palmer, S., Almeida, J., Sims, S., Wilming, L.G., Rogers, J., de Jong, P.J., Carrington, M., Elliott, J.F., Sawcer, S., Todd, J.A., Trowsdale, J. & Beck, S. (2006) Genetic Analysis of Completely Sequenced Disease-Associated MHC Haplotypes Identifies Shuffling of Segments in Recent Human History. *PLoS Genetics*, **2**, e9.

Trowsdale, J. (2011) The MHC, disease and selection. *Immunology Letters*, **137**, 1–8.

Valilou, R.H., Rafat, S.A., Notter, D.R., Shojda, D., Moghaddam, G. & Nematollahi, A. (2015) Fecal egg counts for gastrointestinal nematodes are associated with a polymorphism in the MHC-DRB1 gene in the Iranian Ghezel sheep breed. *Frontiers in Genetics*, **6**, 1–11.

VanLiere, J.M. & Rosenberg, N.A. (2008) Mathematical properties of the measure of linkage disequilibrium. *Theoretical Population Biology*, **74**, 130–137.

Vasoya, D., Law, A., Motta, P., Yu, M., Muwonge, A., Cook, E., Li, X., Bryson, K., MacCallam, A., Sitt, T., Toye, P., Bronsvoort, B., Watson, M., Morrison, W.I. & Connelley, T. (2016) Rapid identification of bovine MHCI haplotypes in genetically divergent cattle populations using next-generation sequencing. *Immunogenetics*, **68**, 765–781.

Wakeland, E.K., Boehme, S., She, J.X., Lu, C.C., McIndoe, R.A., Cheng, I., Ye, Y. & Potts, W.K. (1990) Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. *Immunologic Research*, **9**, 115–122.

Wang, N., Akey, J.M., Zhang, K., Chakraborty, R. & Jin, L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *American journal of human genetics*, 1227–34.

Watkins, D.I., McAdam, S.N., Liu, X., Strang, C.R., Milford, E.L., Levine, C.G., Garber, T.L., Dogon, A.L., Lord, C.I., Ghim, S.H., Troup, G.M., Hughes, A.L. & Letvin, N.L. (1992) New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature*, **357**, 329–333.

Watson, R.L. (2016) *The Causes and Consequences of Immune Variation in a Wild Mammal*. University of Edinburgh.

Watterson, G.A. (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405–417.

Wedekind, C., Walker, M. & Little, T.J. (2005) The Course of Malaria in Mice: Major Histocompatibility Complex (MHC) Effects, but No General MHC Heterozygote Advantage in Single-Strain Infections. *Genetics*, **170**.

Wedekind, C., Walker, M. & Little, T.J. (2006) The separate and combined effects of MHC genotype, parasite clone, and host gender on the course of malaria in mice. *BMC Genetics*, **7**, 55.

Wedekind, C., Walker, M., Portmann, J., Cenni, B., Müller, R. & Binz, T. (2004) MHC-linked susceptibility to a bacterial infection, but no MHC-linked cryptic female choice in whitefish. *Journal of Evolutionary Biology*, **17**, 11–18.

Wegner, K.M. (2008) Historical and contemporary selection of teleost MHC genes: Did we leave the past behind? *Journal of Fish Biology*, **73**, 2110–2132.

Wegner, K.M., Kalbe, M., Kurtz, J., Reusch, T.B.H., Milinski, M., Mathiaswegner, K. & Martinkalbe. (2003a) Parasite Selection for Immunogenetic Optimality. *Science*, **201**, 1343.

Wegner, K.M., Reusch, T.B.H. & Kalbe, M. (2003b) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16**, 224–232.

Weir, B.S. (1996) *Genetic Data Analysis Ii*. Sinauer, Sunderland, Mass.

Westerdahl, H. (2007) Passerine MHC: genetic variation and disease resistance in the wild. *Journal of Ornithology*, **148**, 469.

Westerdahl, H., Asghar, M., Hasselquist, D. & Bensch, S. (2012) Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 577–584.

Westerdahl, H., Hansson, B., Bensch, S. & Hasselquist, D. (2004) Between-year variation of MHC allele frequencies in great reed warblers: Selection or drift? *Journal of Evolutionary Biology*, **17**, 485–492.

Westerdahl, H., Stjernman, M., Råberg, L., Lannefors, M. & Nilsson, J.Å. (2013) MHC-I Affects Infection Intensity but Not Infection Status with a Frequent Avian Malaria Parasite in Blue Tits. *PLoS ONE*, **8**, e72647.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., von Schantz, T. & Bensch, S. (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 1511–1518.

Whittaker, D.J., Dapper, A.L., Peterson, M.P., Atwell, J.W. & Ketterson, E.D. (2012) Maintenance of MHC Class IIB diversity in a recently established songbird population. *Journal of Avian Biology*, **43**, 109–118.

Wilson, K. & Grenfell, B.T. (1997) Generalized linear modelling for parasitologists. *Parasitology Today*, **13**, 33–38.

Wilson, K., Grenfell, B.T., Pilkington, J.G., Boyd, H.E.G. & Gulland, F.M.D. (2004) Parasites and their impact. *Soay Sheep: Dynamics and Selection in an Island Population* (eds T.H. Clutton-Brock & J.M. Pemberton), pp. 113–154. Cambridge University Press, Cambridge.

Wilson, A.J., Pemberton, J.M., Pilkington, J.G., Clutton-Brock, T.H., Coltman, D.W. & Kruuk, L.E.B. (2007) Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evolutionary Ecology*, **21**, 337–356.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E.B. & Nussey, D.H. (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**, 13–26.

Winternitz, J.C., Minchey, S.G., Garamszegi, L.Z., Huang, S., Stephens, P.R. & Altizer, S. (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society*

*of London B: Biological Sciences*, **280**, 20131605.

Winternitz, J.C., Wares, J.P., Yabsley, M.J. & Altizer, S. (2014) Wild cyclic voles maintain high neutral and MHC diversity without strong evidence for parasite-mediated selection. *Evolutionary Ecology*, **28**, 957–975.

Wittzell, H., Bernot, A., Auffray, C. & Zoorob, R. (1999) Concerted Evolution of Two Mhc Class II B Loci in Pheasants and Domestic Chickens. *Mol. Biol. Evol*, **16**, 479–490.

Woelfing, B., Traulsen, A., Milinski, M. & Boehm, T. (2009) Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **364**, 117–128.

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T. & Richardson, D.S. (2010) MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, **19**, 3064–3075.

Worley, K., Gillingham, M.A.., Jensen, P., Kennedy, L.J., Pizzari, T., Kaufman, J. & Richardson, D.S. (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. *Immunogenetics*, **60**, 233–247.

Wright, H. & Ballingall, K.T. (1994) Mapping and characterization of the DQ subregion of the ovine MHC. *Animal Genetics*, **25**, 243–249.

Wright, H., Ballingall, K.T. & Redmond, J. (1994) The DY sub-region of the sheep MHC contains an A/B gene pair. *Immunogenetics*, **40**, 230–234.

Wright, H., Redmond, J. & Ballingall, K.T. (1995) The sheep orthologue of the HLA-DOB gene. *Immunogenetics*, **43**, 76–79.

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, **88**, 76–82.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–49.

Yang, B., Ren, B., Xiang, Z., Yang, J., Yao, H., Garber, P.A. & Li, M. (2014) Major histocompatibility complex and mate choice in the polygynous primate: the Sichuan snub-nosed monkey ( *Rhinopithecus roxellana* ). *Integrative Zoology*, **9**, 598–612.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R. & Flicek, P. (2016) Ensembl 2016. *Nucleic Acids Research*, **44**, D710–D716.

Yewdell, J.W., Norbury, C.C. & Bennink, J.R. (1999) Mechanisms of Exogenous Antigen Presentation by MHC Class I Molecules in Vitro and in Vivo: Implications for Generating CD8+ T Cell Responses to Infectious Agents, Tumors, Transplants, and Vaccines. *Advances in Immunology*, **73**, 1–77.

Zangenberg, G., Huang, M.-M., Arnheim, N. & Erlich, H. (1995) New HLA–DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genetics*, **10**, 407–414.

Zelano, B. & Edwards, S. V. (2002) An MHC component to kin recognition and mate choice in birds: predictions, progress, and prospects. *The American Naturalist*, **160 Suppl**, S225–S237.

Zhai, T., Yang, H.-Q., Zhang, R.-C., Fang, L.-M., Zhong, G.-H. & Fang, S.-G. (2017) Effects of Population Bottleneck and Balancing Selection on the Chinese Alligator Are Revealed by Locus-Specific Characterization of MHC Genes. *Scientific Reports*, **7**, 5549.

Zhang, M. & He, H. (2013) Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (Lasiopodomys brandtii) from Inner Mongolia, China. *BMC Evolutionary Biology*, **13**, 149.

Zhang, L., Wu, Q., Hu, Y., Wu, H. & Wei, F. (2014) Major histocompatibility complex alleles associated with parasite susceptibility in wild giant pandas. *Heredity*, **114**, 85–93.

Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R. & Weir, B.S. (2014) HIBAG—HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal*, **14**, 192–200.

Ziegler, A., Santos, P.S.C., Kellermann, T. & Uchanska-Ziegler, B. (2010) Self/nonself perception, reproduction, and the extended MHC. *Self/Nonself*, **1**, 176–191.

Zinkernagel, R.M. (1979) Associations between major histocompatibility antigens and susceptibility to disease. , **33**, 201–213.

# Appendix A. Studies which associated MHC diversity with either phenotypes or investigated spatial variation within a single species

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | Bank vole | *Myodes glareolus* | Wild | DQA | SSCP Sanger seq | Phenotypes | Virus | No effect of heterozygosity; Specific alleles | (Deter *et al.* 2008) |
| | Bank vole | *Myodes glareolus* | Wild | DRB | 454 NGS | Spatial; Phenotypes | Parasites | Specific alleles associated with parasite burden | (Kloch *et al.* 2010) |
| | Bank vole | *Myodes glareolus* | Wild | DRB | 454 NGS | Phenotypes | Survival, Parasites | Specific alleles associated with increased/reduced susceptibility & survival | (Kloch *et al.* 2013) |
| | Bighorn sheep | *Ovis canadensis* | Wild | | Microsatellite | Spatial | | Neutral processes more important than selection | (Boyce *et al.* 1997) |
| | Bighorn sheep | *Ovis canadensis* | Wild | DRB | SSCP Sanger seq | Spatial | | MHC variation similar to neutral variation; did find dN greater than dS | (Gutierrez-Espeleta *et al.* 2001) |
| | Brandt's voles | *Lasiopodomys brandtii* | Wild | DRB | SSCP | Phenotypes | Parasites | No associated with heterozygosity; specific allele associated with parasite load, allele varied between regions | (Zhang & He 2013) |
| | Brown hares | *Lepus europaeus* | Wild | DQA | SSCP Sanger seq | Temporal | | No evidence of temporal change in frequencies over 8 years | (Campos *et al.* 2011) |
| | Common vampire bat | *Desmodus rotundus* | Wild | DRB | Sanger seq | Spatial | | MHC diversity equal to neutral diversity | (Salmier *et al.* 2016) |
| | Domestic cattle | *Bos* | Experimental | DRB | Microsatellite | Phenotypes | Ticks | Specific alleles associated with tick infestation | (Acosta-Rodríguez *et al.* 2005) |
| | Domestic cattle | *Bos* | Domestic | Class I; DY | RSCA; Sanger seq | Temporal; Phenotypes | Production, Disease, Fertility | Allele frequency changes over time - rare allele increase & reduction in common allele; specific allele associated with phenotypes | (Codner *et al.* 2012) |
| | Domestic sheep | *Ovis aries* | Domestic | Class I; DY | Microsatellite/ DY hybridisation | Phenotypes | FEC | Specific allele associated with FEC | (Buitkamp *et al.* 1996) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | Domestic sheep | *Ovis aries* | Domestic | DQA2 | SSCP | Phenotypes | Bacteria | Specific allele associated with increased susceptibility | (Gelasakis *et al.* 2013) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Microsatellite | Phenotypes | FEC | Specific allele associated with FEC | (Schwaiger *et al.* 1995) |
| | Domestic sheep | *Ovis aries* | Domestic | Class II | RFLP haplotypes | Phenotypes | Parasites | Specific haplotype associated with resistance to cystic echinococcosis | (Shen *et al.* 2014) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Sanger seq | Phenotypes | Parasites, Antibody response | Specific allele associated with FEC | (Stear *et al.* 2005) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | RFLP | Phenotypes | FEC | Specific allele associated with FEC | (Valilou *et al.* 2015) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB | Microsatellite, SSCP | Phenotypes | Parasites, IgG | | (Outteridge *et al.* 1996) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Sanger seq | Phenotypes | Virus; retroviral tumours | Specific allele associated with disease | (Larruskain *et al.* 2010) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Sanger seq | Phenotypes | FEC | Specific allele associated with resistance and susceptibility in Suffolk breed but none in Texel | (Sayers *et al.* 2005) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Sanger seq | Phenotypes | Virus | Specific alleles associated with resistance | (Herrmann-Hoesing *et al.* 2008b) |
| | Domestic sheep | *Ovis aries* | Domestic | DQA1 | Unknown | Phenotypes | FEC | Specific associated with between DQA1*null and strongyles, but not across breeds | (Forrest *et al.* 2010) |
| | Domestic sheep | *Ovis aries* | Domestic | DQA1 | RFLP | Phenotypes | Fungus | Specific associated with resistance to footrot fungus | (Escayg *et al.* 1997) |
| | Domestic sheep | *Ovis aries* | Domestic | DQA1 | Expression Microarray | Phenotypes | FEC | Specific associated with between DQA1*null and strongyles | (Keane *et al.* 2007) |
| | Domestic sheep | *Ovis aries* | Domestic | DRB1 | Microsatellite | Phenotypes | FEC | Specific associated with FEC | (Charon *et al.* 2002) |
| | Domestic sheep | *Ovis aries* | Domestic | DQA2 | Unknown | Phenotypes | Fungus | Specific associated with prevalence of footrot fungus | (Ennen *et al.* 2009) |
| | Domestic sheep | *Ovis aries* | Domestic | Unknown | Microsatellite | Phenotypes | Bacteria | Specific alleles associated with susceptibility to Johne's disease | (Reddacliff *et al.* 2005) |



200

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | European badger | *Meles meles* | Wild | Class I, DRB | RSCA | Phenotypes | Parasites | Specific allele/haplotype associated with prevalence; heterozygosity not associated with coinfection status, but heterozygote advantage against specific infections | (Sin *et al.* 2014) |
| | European rabbit | *Oryctolagus cuniculus* | Wild | DRB | SSCP | Phenotypes | Parasites | Heterozygosity not associated with parasite burden; specific alleles associated with lower burden but only in heterozygous form | (Oppelt *et al.* 2010) |
| | Fat-tailed dwarf lemur | *Cheirogaleus medius* | Wild | DRB | Sanger seq | Phenotypes | Pathogens | Number of supertypes not associated with parasite burden; specific supertype associated with infection, higher nematode diversity and burden | (Schwensow *et al.* 2007) |
| | Fat-tailed dwarf lemur | *Cheirogaleus medius* | Wild | DRB | Sanger seq | Phenotypes | FEC | Specific allele associated with susceptibility to Ascaris | (Schwensow *et al.* 2010) |
| | Forest musk deer | *Moschus berezovskii* | Domestic | DR, DQ | SSCP Sanger seq | Phenotypes | Disease | Specific alleles associated with disease resistance | (Li *et al.* 2014) |
| | Giant panda | *Ailuropoda melanoleuca* | Wild | DRB | Sanger seq | Phenotypes | Parasites | No associated with heterozygosity; specific allele associated with increased susceptibility (but only in one location) | (Zhang *et al.* 2014) |
| | Grey mouse lemur | *Microcebus murinus* | Wild | DRB | SSCP Sanger seq | Phenotypes | FEC | Specific allele associated with susceptibility to Ascaris | (Schwensow *et al.* 2010) |
| | Grey seals | *Halichoerus grypus* | Wild | DQB | SSCP; Sanger seq | Spatial | | MHC differentiation greater than microsatellites | (Cammen *et al.* 2011) |
| | Hairy-footed gerbil | *Gerbillus paeba* | Wild | DRB | Sanger seq | Phenotypes | FEC | Specific allele associated with increased resistance to a single nematode species | (Harf & Sommer 2005) |
| | Humans | *Homo sapiens* | Wild | Haplotype | | Phenotypes | AIDS | Both heterozygosity and specific alleles associated with extended survival and increased tolerance | (Carrington *et al.* 1999) |
| | Humans | *Homo sapiens* | Wild | Haplotype | | Phenotypes | Parasites | Specific allele associated with response to cestode infection | (Godot *et al.* 2000) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | Humans | Homo sapiens | Wild | Haplotype | | Phenotypes | Malaria | Specific haplotypes associated with increased resistance | (Hill 1991) |
| | Humans | *Homo sapiens* | Wild | Haplotype | | Phenotypes | Virus | Heterozygote advantage - heterozygotes less likely to be persistently infected | (Thursz *et al.* 1997) |
| | Long-tailed giant rat | *Leopoldamys sabanus* | Wild | B | SSCP; Sanger seq | Phenotypes | Parasites, Body condition | Number of alleles associated with resistance to helminths and increased body condition | (Lenz *et al.* 2009c) |
| | Malagasy giant jumping rat | *Hypogeomys antimena* | Wild | DQA | SSCP; Sanger seq | Spatial | | MHC differentiation lower than neutral | (Sommer 2003) |
| | Malagasy mouse lemur | *Microcebus murinus* | Wild | DRB | SSCP Sanger seq | Phenotypes | Parasites | Specific alleles associated with variation in parasite load; no effect of heterozygosity | (Schad *et al.* 2005) |
| | Mice | *Mus* | Experimental | Haplotype | MHC congenic mice | Phenotypes | Fitness | No advantage of heterozygosity in the lab, but heterozygosity decreased pup number in females in enclosures | (Ilmonen *et al.* 2007) |
| | Mice | *Mus* | Experimental | Haplotype | MHC congenic mice | Phenotypes | Coinfection, Bacteria, Viruses | Overdominance | (McClelland *et al.* 2003) |
| | Mice | Mus | Experimental | Haplotype | MHC congenic mice | Phenotypes | Bacteria | MHC heterozygotes had prolonged survival and were better able to resolve Salmonella infection, but dominance not overdominance; MHC type affected pathogen load for secondary infection | (Penn *et al.* 2002) |
| | Mice | *Mus* | Experimental | Haplotype | MHC congenic mice | Phenotypes | Malaria | No heterozygote advantage | (Wedekind, Walker & Little 2005) |
| | Mice | *Mus* | Experimental | Haplotype | MHC congenic mice | Phenotypes | Malaria | *Specific haplotype associated with course of malaria* | (Wedekind, Walker & Little 2006) |
| | Mice & Humans | *Mus & Homo sapiens* | Experimental | Haplotype | MHC congenic mice | Phenotypes | Leishmanial infections | Associated with haplotypes with susceptibility | Blackman1996 |


202

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | Montane voles | *Microtus montanus* | Wild | DRB | 454 NGS | Phenotypes | Parasites | No association with parasite diversity or prevalence; specific allele associated with reduced cestode burden | (Winternitz *et al.* 2014) |
| | Mountain brushtail possums | *Trichosurus cunninghami* | Wild | DAB1 | Microsatellite | Phenotypes | Survival, FEC | He zygotes had reduced FEC | (Banks *et al.* 2010) |
| | New Zealand sea lion | *Phocarctos hookeri* | Wild | DRB, DQA | Sanger seq | Phenotypes | Survival | Heterozygote advantage and specific alleles associated with survival | (Osborne *et al.* 2015) |
| | Pallas's mastiff bat | *Molossus molossus* | Wild | DRB | Sanger seq | Spatial | | MHC diversity greater than neutral diversity | (Salmier *et al.* 2016) |
| | Plains zebra | *Equus quagga* | Wild | DRA | Sanger seq | Phenotypes | Endoparasites, Ectoparasites | Rare alleles associated with increased GI parasites, common alleles increased tick burden; heterozygote advantage amongst common alleles | (Kamath *et al.* 2014) |
| | Racoon | *Procyon lotor* | Wild | DRB | Sanger seq | Spatial | | MHC differentiation greater than microsatellites | (Kyle *et al.* 2014) |
| | Racoon | *Procyon lotor* | Semi-experimental | DRB | 454 NGS | Phenotypes | IgG, IgM | Specific allele/supertype associated with IgG for canine distemper virus, and another allele for IgG for parvovirus | (Ruiz-López *et al.* 2014) |
| | San Nicolas Island fox | *Urocyon littoralis* | Wild | DRB | SSCP, microsatellite | Spatial | | Lower differentiation at MHC than microsatellites - Circumstances of bottleneck required to have maintained the MHC diversity are unrealistic. | (Aguilar *et al.* 2004) |
| | Seba's short-tailed bat | *Carollia perspicillata* | Wild | DRB | Sanger seq | Spatial | | MHC diversity equal to neutral diversity | (Salmier *et al.* 2016) |
| | Soay sheep | *Ovis aries* | Wild | DRB | Microsatellite | Spatial; Temporal | | Spatial MHC variation less than microsatellites in some years; Some evidence of temporal variation | (Charbonnel & Pemberton 2005) |
| | Soay sheep | *Ovis aries* | Wild | DRB | Microsatellite | Phenotypes | Survival, FEC | Specific alleles associated with FEC and survival | (Paterson 1998) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | Southeast Asian house mouse | *Mus musculus castaneus* | Wild | Unknown | Microsatellite | Spatial | | Neutral processes stronger than selection | (Huang & Yu 2003) |
| | Spotted suslik | *Spermophilus suslicus* | Wild | DRB | SSCP Sanger seq | Phenotypes | Blood, parasites | Heterozygosity not associated with parasite prevalence or infection; specific allele associated with parasite; deviations from HWE | (Biedrzycka *et al.* 2011) |
| | Striped mouse | *Rhabdomys pumiloo* | Wild | DRB | SSCP | Phenotypes | Parasites | Heterozygote advantage via divergent allele advantage for infection intensity; specific alleles associated with parasite burden | (Froeschke & Sommer 2012) |
| | Striped mouse | *Rhabdomys pumiloo* | Wild | DRB | SSCP | Spatial | | MHC differentiation less than microsatellites | (Froeschke & Sommer 2014) |
| | Striped mouse | *Rhabdomys pumiloo* | Wild | DRB | SSCP Sanger seq | Phenotypes | FEC | Heterozygosity not associated with infection status or FEC; Specific allele associated with FEC | (Froeschke & Sommer 2005) |
| | Talas tuco-tuco | *Ctenomys talarum* | Wild | DRB | Sanger seq | Phenotypes | Parasites, Antibody response | Excess of heterozygotes; specific alleles; no association between heterozygosity with phenotypes | (Cutrera *et al.* 2011) |
| | Water vole | *Arvicola terrestris* | Wild | DRB1, DQA1 | Sanger seq | Spatial | | Density-related changes - local selection in fragmented pops, then homogenization | (Bryja *et al.* 2007) |
| | Water vole | *Arvicola terrestris* | Wild | DRB | SSCP | Temporal | | MHC allele frequencies maintained across a bottleneck, probably due to MHC-dependent survival | (Oliver & Piertney 2012) |
| | Water vole | *Arvicola terrestris* | Wild | DRB | Sanger seq | Spatial | | Spatio-temporal variation - genetic drift, not selection, over time; drift/directional selection over space | (Oliver *et al.* 2009a) |
| | Water vole | *Arvicola terrestris* | Wild | DRB | SSCP | Phenotypes | Multiple parasites | Specific genotypes associated with parasite burden; heterozygosity associated with coinfection by fewer parasites | (Oliver *et al.* 2009b) |
| | Water vole | *Arvicola scherman* | Wild | DRB, DQA1 | SSCP Sanger seq | Spatial; Phenotypes | Parasites | differentiation MHC less than microsatellites; specific alleles and nematode presence | (Tollenaere *et al.* 2008) |



204

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Mammal | White-tailed deer | *Odocoileus virginianus* | Wild | DRB | Sanger seq | Phenotypes | Parasites | Possible overdominance; associations with nematodes and ticks | (Ditchkoff *et al.* 2005) |
| | Wolves | *Canis lupus* | Wild | DRB1, DQA1, DQB1 | Sanger seq | Spatial; Phenotypes | Parasites | MHC differentiation less than microsatellites; heterozygosity associated with reduced infection; specific alleles associated with infection | (Niskanen *et al.* 2014) |
| | Yellow-necked mouse | *Apodemus flavicollis* | Wild | DRB | SSCP Sanger seq | Phenotypes | FEC | Specific alleles affected susceptibility to nematodes | (Meyer-Lucht & Sommer 2005) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Fish | Arctic charr | *Salvelinus alpinus* | Wild | Class II B | DGGE | Phenotypes | Parasites | Heterozygosity associated with reduced parasite burden, increased body condition and increased survival | (Kekäläinen *et al.* 2009) |
| | Atlantic salmon | *Salmo salar* | Wild | Class I, Class II | Microsatellites | Spatial; Temporal | | MHC differentiation greater than neutral | (Consuegra *et al.* 2011) |
| | Atlantic salmon | *Salmo salar* | Wild | Class II B | DGGE | Spatial | | MHC diversity greater than microsatellites | (Dionne *et al.* 2007) |
| | Atlantic salmon | *Salmo salar* | Wild | Class II B | DGGE | Spatial; Phenotypes | Bacteria | Specific alleles associated with pathogen resistance | (Dionne *et al.* 2009) |
| | Atlantic salmon | *Salmo salar* | Experimental | Class I, Class II | Microsatellites | Phenotypes | Disease | Specific allele associated with resistance/susceptibility | (Grimholt *et al.* 2003) |
| | Atlantic salmon | *Salmo salar* | Experimental | Class I, Class II | Microsatellites | Phenotypes | Disease | Specific allele associated with resistance/susceptibility | (Kjøglum *et al.* 2008) |
| | Atlantic salmon | *Salmo salar* | Wild | Class II B | SSCP | Spatial | | Neutral processes more important than selection | (Landry & Bernatchez 2001) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Fish | Atlantic salmon | *Salmo salar* | Wild | Class II B | RFLP | Spatial | | More pronounced geographical structure in MHC genes than in neutral markers | (Langefors 2005) |
| | Atlantic salmon | *Salmo salar* | Experimental | Class II B | DGGE | Phenotypes | Bacteria | Specific alleles associated with resistance | (Langefors *et al.* 2001a) |
| | Atlantic salmon | *Salmo salar* | Experimental | Class II B | DGGE | Phenotypes | Parasites | Specific alleles associated with resistance/susceptibility; no effect of heterozygosity on survival | (Lohm *et al.* 2002) |
| | Blunt-head cichlid | Tropheus moorii | Wild | Class II B | Sanger seq | Phenotypes | Parasites, Fat deposits | Intermediate allele number associated with highest perivisceral fat deposits but not parasites | (Hablützel *et al.* 2014) |
| | Brook charr | *Salvelinus fontinalis* | Experimental | Class II B | SSCP; Sanger seq | Phenotypes | Disease | Heterozygosity reduced resistance, specific allele associated with increased resistance | (Croisetière *et al.* 2008) |
| | Brown trout | *Salmo trutta* | Wild | Class II B | SSCP; Sanger seq | Spatial | | MHC differentiation equal to neutral variation | (Campos, Posada & Morán 2006) |
| | Brown trout | *Salmo trutta* | Wild | | Microsatellites | Spatial; Temporal | | Fluctuation in temporal selection | (Hansen *et al.* 2007) |
| | Brown trout | *Salmo trutta* | Wild | Class II B | Microsatellite | Spatial; Temporal | | Spatial and temporal variation | (Jensen *et al.* 2008) |
| | California costal steelhead | *Oncorhynchus mykiss* | Wild | Class II B | SSCP | Spatial | | Weak geographical structure in MHC diversity | (Aguilar & Garza 2006) |
| | Chinook salmon | *Oncorhynchus tshawytscha* | Experimental | Class II B | DGGE | Phenotypes | Virus, Bacteria, Parasites | Heterozygosity increased survival | (Arkush *et al.* 2002) |
| | Chinook salmon | *Oncorhynchus tshawytscha* | Experimental | Class I, Class II | SSCP | Spatial; Phenotypes | Survival | Class I nucleotide diversity associated with survival; Class II diversity disadvantageous in one pop; local adaptation | (Evans, Neff & Heath 2010a) |
| | Chinook salmon | *Oncorhynchus tshawytscha* | Wild | Class I, Class II | SSCP | Spatial | | Class I differentiation greater than microsatellites; CII differentiation less than microsatellites | (Evans, Neff & Heath 2010b) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Fish | European chub | Squalius cephalus | Wild | DAB | SSCP | Phenotypes | Parasites | DAB1 specific alleles with ectoparasites, DQB3, with endoparasites | (Seifertová *et al.* 2016) |
| | European minnows | *Phoxinus phoxinus* | Wild | DAB | 454 NGS | Spatial; Phenotypes | Parasites | MHC differentiation greater than microsatellite; specific allele associated with bacterial diversity | (Collin *et al.* 2013) |
| | Gila topminnow | *Poeciliopsis o. occidentalis* | Wild | Class II B | Sanger seq SSCP | Spatial | | Neutral processes stronger than selection | (Hedrick, Parker & Lee 2001) |
| | Lake whitefish | *Coregonus clupeaformis* | Wild | Class II B | 454 NGS | Phenotypes | Parasites | Increased allele number associated with increased infection; specific alleles associated with clusters of pathogens | (Pavey *et al.* 2013) |
| | Mangrove rivulus | *Kryptolebias marmoratus* | Wild | Class I | 454 NGS | Phenotypes | Parasites | MHC diversity associated with parasite load; non-random loss of MHC alleles despite self-fertilisation | (Ellison *et al.* 2012) |
| | Shortfin molly | *Poecilia mexicana* | Wild | Class II B | Sanger seq | Spatial | | Increased parasite load within community correlated with increased MHC diversity, but MHC alleles very similar between communities | (Tobler *et al.* 2014) |
| | Sockeye salmon | *Oncorhynchus nerka* | Wild | Class II B | DGGE | Spatial; Temporal | | More pronounced geographical structure at MHC than neutral markers; temporal variation | (Miller *et al.* 2001) |
| | Sockeye salmon | *Oncorhynchus nerka* | Wild | Class II B | DGGE | Spatial | | MHC differentiation greater than neutral | (Miller *et al.* 2001) |
| | Sticklebacks | *Gasterosteus aculeatus* | Wild | Class II B | RSCA | Spatial | | Differentiation in MHC between habitats | (Eizaguirre *et al.* 2011) |
| | Sticklebacks | *Gasterosteus aculeatus* | Experimental | Class II B | RSCA | Spatial; Phenotypes | Parasites | Differentiation in MHC associated with parasite reduction | (Eizaguirre *et al.* 2012a) |
| | Sticklebacks | *Gasterosteus aculeatus* | Experimental | Class II B | RSCA | Phenotypes | Parasites | Only resistant genotypes increased in next generation | (Eizaguirre *et al.* 2012b) |
| | Sticklebacks | *Gasterosteus aculeatus* | Experimental | Class II B | SSCP | Phenotypes | Lifetime reproductive success | Intermediate number of alleles associated with highest LRS | (Kalbe *et al.* 2009) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Fish | Sticklebacks | *Gasterosteus aculeatus* | Experimental | Class II B | SSCP | Phenotypes | Parasites | Increased MHC diversity reduced tapeworm infection burden and mass, and reduction in probability of microsporidian infection | (Kurtz *et al.* 2004) |
| | Sticklebacks | *Gasterosteus aculeatus* | Experimental | Class II B | RSCA | Phenotypes | Immune assay, Parasites, Reproductive success | Immune measure increased with MHC distance | (Lenz *et al.* 2009b) |
| | Sticklebacks | *Gasterosteus aculeatus* | Wild | Class II B | SSCP; Sanger seq | Phenotypes | Parasites | Intermediate number of alleles associated with lowest parasite load | (McCairns *et al.* 2011) |
| | Sticklebacks | *Gasterosteus aculeatus* | Experimental | | Sanger seq mRNA | Phenotypes | Parasites | Intermediate diversity reduced parasite load, but only in co-infection | (Wegner *et al.* 2003a) |
| | Sticklebacks | *Gasterosteus aculeatus* | Wild | Class II B | SSCP | Phenotypes | Parasites | Intermediate diversity and suggest overdominance | (Wegner *et al.* 2003b) |
| | Sticklebacks | *Gasterosteus aculeatus* | Semi-experimental | Class II B | SSCP | Phenotypes | Parasites, Survival | Intermediate number of alleles survived and had lowest parasites. | (Wegner 2008) |
| | Sticklebacks | *Gasterosteus aculeatus* | Semi-experimental | Class II B | 454 NGS | Phenotypes | Survival | Survival highest in locally rare genotypes | (Bolnick & Stutz 2017) |
| | Trinidadian guppy | *Poecilia reticulata* | Wild | Class II B | SSCP | Spatial; Phenotypes | Parasites | Specific allele increased dosage associated with reduction in parasite load | (Fraser & Neff 2010) |
| | Trinidadian guppy | *Poecilia reticulata* | Wild | Class II B | Sanger seq | Spatial | | MHC differentiation less than neutral | (Fraser, Ramnarine & Neff 2010) |
| | Trinidadian guppy | *Poecilia reticulata* | Wild | Class II B | Sanger seq | Spatial | | MHC differentiation less than neutral | (van Oosterhout *et al.* 2006) |
| | Whitefish | Coregonus sp. | Experimental | Class II B | SSCP | Phenotypes | Bacteria | Specific allele associated with survival | (Wedekind *et al.* 2004) |



208

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Birds | Lesser kestrel | *Falco naumanni* | Wild | Class II B | Sanger seq | Spatial | | More pronounced geographical structure in MHC genes than in neutral markers | (Alcaide *et al.* 2008) |
| | Attwater's Prairie-chicken | *Tympanuchus cupido attwateri* | Semi-natural | Class I, Class II | 454 NGS | Phenotypes | Survival | Specific alleles associated with survival | (Bateson *et al.* 2016) |
| | House sparrows | *Passer domesticus* | Wild | Class I | CE-SSCP; Sanger seq | Spatial | | Differentiation at MHC less than neutral | (Bichet *et al.* 2015) |
| | House sparrows | *Passer domesticus* | Experimental | Class I | DGGE | Phenotypes | Nestling survival | Increased allelic diversity associated with increased nestling survival | (Bonneaud *et al.* 2004) |
| | House sparrows | *Passer domesticus* | Experimental | Class I | DGGE | Phenotypes | Antibody response | Allele number not associated with response to antigens; specific allele associated with response to antigens | (Bonneaud *et al.* 2005) |
| | House sparrows | *Passer domesticus* | Wild | Class I | DGGE | Phenotypes | Malaria | Specific allele associated with resistance/susceptibility; spatial variation | (Bonneaud *et al.* 2006) |
| | Seychelles warbler | *Acrocephalus sechellensis* | Wild | Class I | RSCA | Phenotypes | Survival | Allele number associated with increased juvenile survival; specific allele associated with lifespan | (Brouwer *et al.* 2010) |
| | Yellowthroats | *Geothlypis trichas* | Wild | Class I, Class II | 454 NGS | Phenotypes | Survival, Malaria | Allele number associated with increased survival; specific allele associated with resistance to malaria | (Dunn *et al.* 2013) |
| | Great snipe | *Gallinago media* | Wild | Class II B | Sanger seq | Spatial | | Selection not neutral processes | (Ekblom *et al.* 2007) |
| | Berthelot's pipet | *Anthus berthelotii* | Wild | Class I | 454 NGS | Spatial | | MHC variation equal to neutral; specific alleles associated with malaria risk and distance from poultry farm | (González-Quevedo *et al.* 2016) |
| | House finch | *Carpodacus mexicanus* | Wild | Class II | SSCP | Phenotypes | Bacteria | Intermediate to high MHC number more resistant; MHC diversity increased in population after epidemic but not neutral | (Hawley & Fleischer 2012) |
| | Golden pheasant | *Chrysolophus pictus* | Wild | Class II B | SSCP | Spatial | | MHC differentiation less than neutral | (He *et al.* 2017) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Birds | Rufus-collared sparrows | *Zonotrichia capensis* | Wild | Class I | 454 NGS | Spatial; Phenotypes | Malaria | Spatial variation in MHC diversity, allele number, proteins, supertypes; higher diversity associated with higher prevalence of malaria | (Jones *et al.* 2014) |
| | House sparrows | *Passer domesticus* | Wild | Class I | RSCA | Phenotypes | Survival, Recruitment | No overall association with survival or recruitment | (Karlsson *et al.* 2015) |
| | Chinese egret | *Egretta eulophotes* | Wild | DAB | SSCP | Spatial | | MHC differentiation between populations | (Lei *et al.* 2015) |
| | House sparrows | *Passer domesticus* | Wild | Class I | DGGE | Phenotypes | Malaria, Parasites | Specific allele associated with increased infection with malaria and reduction in blood parasite | (Loiseau *et al.* 2008) |
| | House sparrows | *Passer domesticus* | Wild | Class I | DGGE | Spatial | | Differentiation at MHC greater than microsatellites | (Loiseau *et al.* 2009) |
| | House sparrows | *Passer domesticus* | Wild | Class I | DGGE | Spatial; Phenotypes | Infection | Location by allele interaction - variation amongst sites in associations between specific allele and infection status | (Loiseau *et al.* 2011) |
| | House sparrows | *Passer domesticus* | Semi-natural | Class I | 454 NGS | Phenotypes | Offspring survival, Body mass, Size | Specific MHC functional group associated with reduced survival, body mass and size | (Lukasch *et al.* 2017) |
| | Chicken | *Gallus gallus* | Experimental | Haplotypes | MHC congenic lines | Phenotypes | Virus | Not heterozygote advantage; specific haplotypes associated with antibody response | (Martin *et al.* 1989) |
| | South Island robin | *Petroica australis australis* | Wild | Class II | SSCP | Spatial | | Neutral processes greater than selection | (Miller & Lambert 2004b) |
| | Chicken | *Gallus gallus* | Experimental | Haplotypes | MHC congenic lines | Phenotypes | Parasites | Specific alleles associated with resistance | (Owen, Delany & Mullens 2008) |
| | Collared flycatchers | *Ficedula albicollis* | Wild | Class I | 454 NGS | Phenotypes | Parasites, Lifetime reproductive success | No associations between diversity and prevalence of blood parasite, nor LRS | (Radwan *et al.* 2012) |
| | Great tits | *Parus major* | Wild | Class I | 454 NGS | Phenotypes | Fitness | Supertypes associated with survival, | (Sepil *et al.* 2012) |



210

| | | | | | | | | annual recruitment and LRS | |
|---|---|---|---|---|---|---|---|---|---|
| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
| Bird | Great tits | *Parus major* | Wild | Class I | 454 NGS | Phenotypes | Malaria | Different supertypes associated with qualitative and quantitative resistance | (Sepil *et al.* 2013b) |
| | Great reed warblers | *Acrocephalus arundinaceus* | Wild | Class I | DGGE | Temporal | | Variation between cohorts at MHC greater than neutral | (Westerdahl *et al.* 2004) |
| | Great reed warblers | *Acrocephalus arundinaceus* | Wild | Class I | DGGE | Phenotypes | Malaria | Increased diversity associated with increased prevalence | (Westerdahl *et al.* 2005) |
| | Great reed warblers | *Acrocephalus arundinaceus* | Wild | Class I | DGGE | Phenotypes | Malaria | Specific allele associated with quantitative resistance | (Westerdahl *et al.* 2012) |
| | Blue tits | *Cyanistes caeruleus* | Wild | Class I | RSCA | Phenotypes | Malaria | Specific alleles associated with infection intensity but not clearance | (Westerdahl *et al.* 2013) |
| | Red junglefowl | *Gallus gallus* | Semi-natural | Class II B | RSCA | Phenotypes | Endoparasite infection | Heterozygotes survived infection longer but due to reduced effect of susceptible haplotype - not overdominance | (Worley *et al.* 2010) |



| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Amphibian | Lowland leopard frog | *Lithobates yavapaiensis* | Experimental | Class II B | Sanger seq | Phenotypes | Fungus | Specific alleles & heterozygosity | (Savage & Zamudio 2011) |
| Reptile | Chinese alligator | Alligator sinensis | Wild | Class I, Class II | SSCP | Spatial | | MHC differentiation less than neutral | (Zhai *et al.* 2017) |
| Reptile | Gopher tortoise | Gopherus polyphemus | Wild | Class I, Class II | Sanger seq | Spatial | | Natural processes generally more important (threatened population) | (Elbers, Clostio & Taylor 2017) |
| Reptile | Ornate dragon lizard | *Ctenophorus ornatus* | Wild | Class I | 454 NGS | Spatial; Phenotypes | Ectoparasites | No spatial differentiation; diversity not associated with tick burden; number of alleles positively associated with tick burden in reserves | (Radwan *et al.* 2014) |
| Reptile | Sand lizard | *Lacerta agilis* | Wild | Class I | RFLP | Phenotypes | Ectoparasites | Specific allele associated with reduced ectoparasites | (Olsson *et al.* 2005) |
| Reptile | Tuatara | *Sphenodon spp.* | Wild | Class I | DGGE | Spatial | | MHC differentiation greater than neutral | (Miller, Allendorf & Daugherty 2010) |

| Taxon | Species | Binomial | Wild or experimental | Loci | Marker type | Main aspect of study | Trait(s) | Main conclusion | Refs |
|---|---|---|---|---|---|---|---|---|---|
| Reptile | Water python | *Liasis fuscus* | Wild | | RFLP | Phenotypes | Parasites | Intermediate fragment number decreased parasite number; specific frags associated with resistance | (Madsen & Ujvari 2006) |
| Reptile | Eastern Massasauga | *Sistrurus catenatus* | Wild | Class II B | 454 NGS | Spatial | | Variation between populations suggestive of neutral processes but does not rule out selective processes | (Jaeger *et al.* 2016) |



212

# Appendix B. MHC allelic nomenclature

The International Polymorphism Database (IPD) is a centralised repository for polymorphic immune genes, and includes the IPD – MHC (http://www.ebi.ac.uk/ipd/mhc) for MHC genes of various species. MHC alleles from sheep (*Ovis*) are collated by Dr. Keith Ballingall of the Moredun Research Institute for the IPD – MHC. Nomenclature for novel alleles is assigned following standardised format (Marsh *et al.* 2010) and described at

https://www.ebi.ac.uk/ipd/mhc/group/OLA/page/nomenclature. This nomenclature can be described using the example allele: *Ovar-DRB1*01:02:03*.

*Ovar-DRB1* indicates the species and locus.

The first two digits (e.g. *Ovar-DQB1***01**:02:03*) represent the allelic family. Within this family, alleles can differ by a maximum of four amino acids within exon 2, and up to an additional four amino acids throughout the remaining transcript.

The second two digits (e.g. *Ovar-DQB1*01:**02**:03*) refer to a coding change within the allelic family.

The final two digits (e.g. *Ovar-DQB1*01:02:**03*) indicate synonymous substitutions.

# Appendix C. Comparison of OLADRB microsatellite to class IIa haplotypes

The OLADRB microsatellite, which is located in intron 2 of the class IIa *DRB1* locus, was genotyped by Paterson *et al.* (1998) in individuals born between 1985 and 1994, and used in association studies between lamb and juvenile faecal egg counts and survival. Paterson (1998) sequenced the six most common OLADRB alleles, out of a total of eight alleles, and Table C.1 shows how the OLADRB allelic sequences match the *DRB1* alleles identified in the Soay sheep in Chapter 2. All six sequenced OLADRB alleles matched *DRB1* alleles, and, like the *DRB1* alleles, did not separate haplotypes A and G, nor F and H.

Table C.1. OLADRB allelic sequences matched to *DRB1* exon 2 sequences.

| OLADRB allele | GenBank accession no. | DRB1 allele | IPD accession no[a] | Class IIa haplotype |
|---|---|---|---|---|
| 205 | Y10245 | *22:01 | IPD0008646 | F and H |
| 213 | Y10246 | *13:01 | IPD0008616 | E |
| 257 | Y10247 | *01:02 | IPD0008612 | B |
| 263 | Y10248 | *01:01 | IPD0008590 | A and G |
| 267 | Y10249 | *01:01 | IPD0008590 | A and G |
| 269[b] | *NA* | | | |
| 276 | Y10250 | *03:02 | IPD0008588 | C |
| 287[b] | *NA* | | | |

[a] From IPD-MHC database at http://www.ebi.ac.uk/ipd/mhc/
[b] Rare allele not sequenced

After generating diplotype data for large numbers of individuals in Chapter 4, 941 individuals that had previously genotyped at the OLADRB microsatellite by Paterson *et al.* (1998) also now had class IIa diplotypes. Figure C.0.1 shows the comparison of OLADRB genotypes with class IIa diplotypes. Many OLADRB genotypes were found to correspond to multiple class IIa diplotypes. Class IIa haplotypes B, C and E are each uniquely associated with a single *DRB1* allele (see Chapter 2 and Table C.1), yet Figure C.0.1 shows that although the corresponding homozygous OLADRB genotypes (257, 276 and 213, respectively) are associated with multiple class IIa diplotypes. This suggests there may be a large number of genotyping errors in the OLADRB genotypes, which is supported by a high number of parent-offspring mismatches in the Soay sheep database for this locus (J. Pemberton, pers. comm.). Alternatively, it is possible that the OLADRB microsatellites evolve more rapidly than the class IIa genes, though it seems

unlikely the level of variation observed in Figure C.0.1 would be possible given the location of the microsatellite within intron 2 of *DRB1*. The main associations between OLADRB alleles and class IIa haplotypes are shown in Table C.2. This data clearly shows that the OLADRB microsatellite was not able to sufficiently represent the variation across the class IIa haplotypes, which may have been exacerbated by genotyping errors.

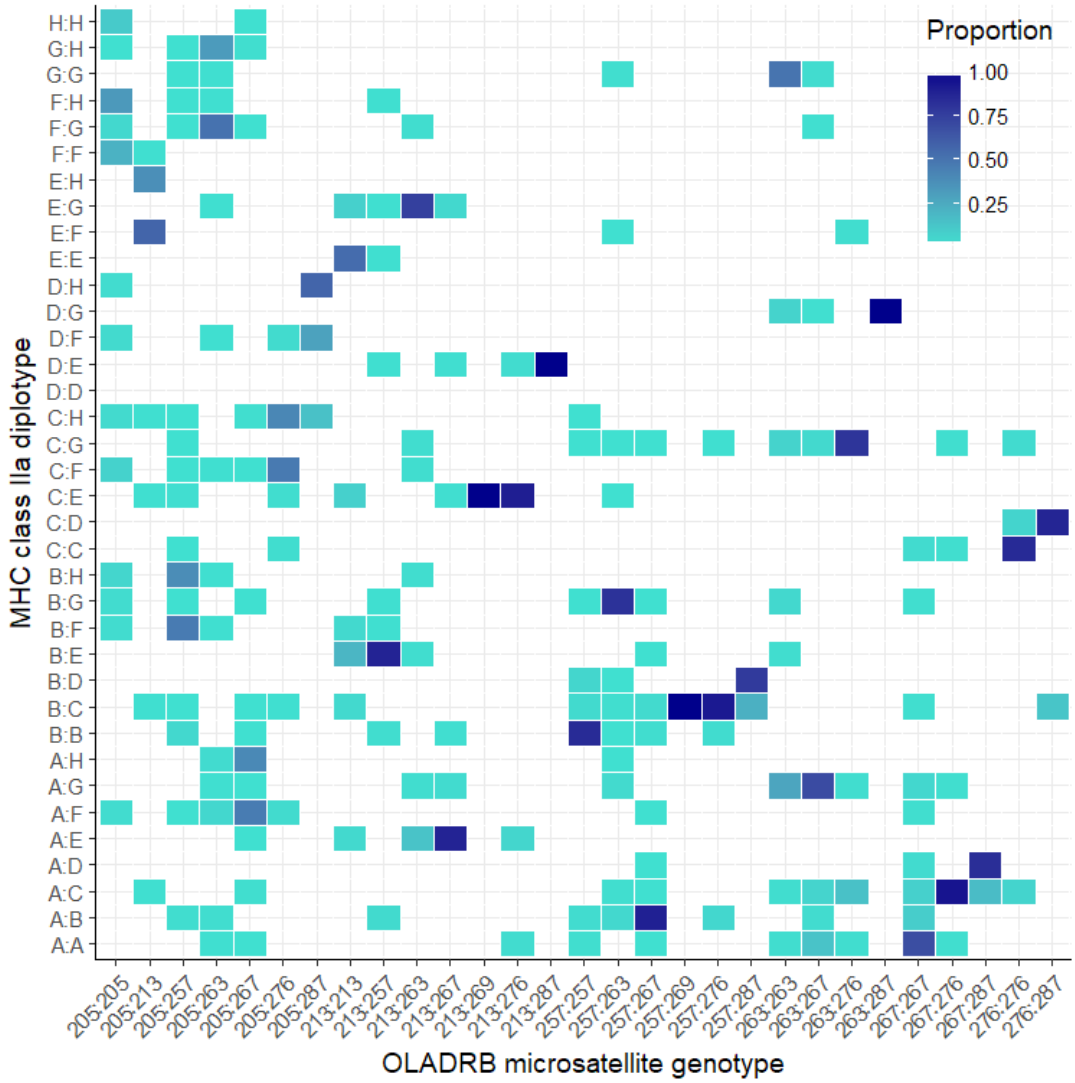



Figure C.0.1. Comparison of OLADRB microsatellite genotypes (allele size in bp) from Paterson *et al.* (1998) with MHC class IIa diplotypes genotyped by KASP in Chapter 4. The intensity of the blue colour of the square represents the proportion of individuals carrying a particular OLADRB genotype that also had the corresponding class IIa diplotype.

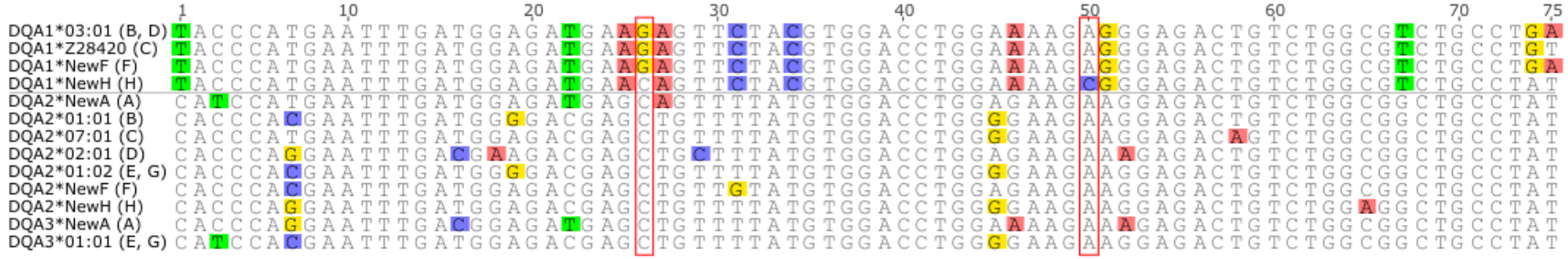
Table C.2. Main associations between OLADRB alleles and class IIa haplotypes.

| OLADRB allele | DRB1 allele | Predominant class IIa haplotype | Other frequently associated haplotypes |
|---|---|---|---|
| 205 | *22:01 | F and H | |
| 213 | *13:01 | E | |
| 257 | *01:02 | B | |
| 263 | *01:01 | G | A |
| 267 | *01:01 | A | G |
| 269[a] | | | |
| 276 | *03:02 | C | |
| 287[b] | *10:01 | D | |

[a] Only three individuals were identified, thus class IIa haplotype could not be assigned

# Appendix D. DQA1 sequences

Aligned nucleotide sequences of all DQA alleles identified in the Soay sheep study population. DQA1 SNPs genotyped by KASP are indicated by red boxes. High sequence similarity between *DQA1* and the *DQA2/2-like* alleles, suggests cross-amplification may be possible.



# Appendix E. HWE tests for each cohort of four LH stages

| | Known conceived | | | Live Born | | |
|---|---|---|---|---|---|---|
| | | Exact test | Heterozygote excess | | Exact test | Heterozygote excess |
| Birth Year | n | P value | P value | n | P value | P value |
| 1985 | 6 | 0.4088 | 0.1268 | 6 | 0.4072 | 0.8056 |
| 1986 | 37 | 0.2795 | 0.7481 | 37 | 0.2707 | 0.7513 |
| 1987 | 80 | 0.4882 | 0.5918 | 80 | 0.4795 | 0.572 |
| 1988 | 30 | 0.5776 | 0.1158 | 30 | 0.5779 | 0.1162 |
| 1989 | 94 | 0.1863 | 0.8455 | 94 | 0.0917 | 0.9088 |
| 1990 | 121 | 0.978 | 0.2055 | 121 | 0.9775 | 0.2041 |
| 1991 | 177 | 0.9477 | 0.0848 | 177 | 0.9301 | 0.1025 |
| 1992 | 172 | 0.4599 | 0.181 | 172 | 0.5481 | 0.1988 |
| 1993 | 197 | 0.8644 | 0.64 | 197 | 0.872 | 0.6357 |
| 1994 | 186 | 0.6854 | 0.7723 | 186 | 0.7467 | 0.7897 |
| 1995 | 184 | 0.3743 | 0.0906 | 184 | 0.564 | 0.1261 |
| 1996 | 207 | 0.1353 | 0.8939 | 207 | 0.1957 | 0.8683 |
| 1997 | 218 | 0.4091 | 0.4227 | 218 | 0.4027 | 0.421 |
| 1998 | 226 | 0.6183 | 0.9772 | 226 | 0.6257 | 0.975 |
| 1999 | 162 | 0.7191 | 0.8833 | 162 | 0.764 | 0.8866 |
| 2000 | 189 | 0.2599 | 0.6619 | 189 | 0.2409 | 0.659 |
| 2001 | 249 | 0.7479 | 0.5833 | 249 | 0.6969 | 0.5221 |
| 2002 | 141 | 0.1273 | 0.7734 | 141 | 0.1218 | 0.7791 |
| 2003 | 234 | 0.1714 | 0.9879 | 234 | 0.1802 | 0.9884 |
| 2004 | 303 | 0.177 | 0.0265 | 303 | 0.1555 | 0.0206 |
| 2005 | 319 | 0.8388 | 0.1847 | 319 | 0.5457 | 0.1891 |
| 2006 | 220 | 0.0952 | 0.3741 | 220 | 0.0807 | 0.191 |
| 2007 | 263 | 0.5154 | 0.0236 | 263 | 0.3793 | 0.0626 |
| 2008 | 250 | 0.2011 | 0.1926 | 250 | 0.2596 | 0.2415 |
| 2009 | 241 | 0.9721 | 0.2726 | 241 | 0.9894 | 0.2786 |
| 2010 | 42 | 0.1755 | 0.3008 | 42 | 0.9125 | 0.2337 |
| 2011 | 331 | 0.2125 | 0.4636 | 331 | 0.4124 | 0.5648 |
| 2012 | 279 | 0.0232 | 0.9599 | 279 | 0.2627 | 0.8421 |

Appendix E

| | Survived to 1st August | | | Survived to 2nd August | | |
|---|---|---|---|---|---|---|
| Birth Year | n | Exact test P value | Heterozygote excess P value | n | Exact test P value | Heterozygote excess P value |
| 1985 | 6 | 0.4072 | 0.8056 | 5 | 0.6194 | 0.7916 |
| 1986 | 37 | 0.2707 | 0.7513 | 35 | 0.3797 | 0.7874 |
| 1987 | 80 | 0.4795 | 0.572 | 67 | 0.752 | 0.6707 |
| 1988 | 30 | 0.5779 | 0.1162 | 16 | 0.0633 | 0.328 |
| 1989 | 68 | 0.0329 | 0.7467 | 51 | 0.1241 | 0.7553 |
| 1990 | 116 | 0.9875 | 0.1478 | 100 | 0.9864 | 0.2284 |
| 1991 | 160 | 0.9136 | 0.207 | 42 | 0.3631 | 0.5565 |
| 1992 | 97 | 0.1244 | 0.3845 | 75 | 0.0823 | 0.0832 |
| 1993 | 168 | 0.9257 | 0.5961 | 71 | 0.9794 | 0.2035 |
| 1994 | 133 | 0.7287 | 0.6986 | 23 | 0.3413 | 0.4374 |
| 1995 | 137 | 0.8982 | 0.0649 | 120 | 0.8658 | 0.1276 |
| 1996 | 185 | 0.0958 | 0.9039 | 64 | 0.5661 | 0.3911 |
| 1997 | 189 | 0.1665 | 0.1659 | 59 | 0.546 | 0.3651 |
| 1998 | 202 | 0.2708 | 0.9932 | 14 | 0.2176 | 0.962 |
| 1999 | 132 | 0.8812 | 0.9161 | 106 | 0.7093 | 0.9633 |
| 2000 | 161 | 0.8449 | 0.5364 | 128 | 0.7988 | 0.4579 |
| 2001 | 245 | 0.6996 | 0.5316 | 15 | 0.0113 | 0.676 |
| 2002 | 131 | 0.1908 | 0.7603 | 117 | 0.2018 | 0.6822 |
| 2003 | 211 | 0.106 | 0.9951 | 171 | 0.1054 | 0.9446 |
| 2004 | 210 | 0.0174 | 0.037 | 13 | 0.9731 | 0.3029 |
| 2005 | 135 | 0.7427 | 0.2111 | 62 | 0.947 | 0.0352 |
| 2006 | 166 | 0.0962 | 0.2103 | 57 | 0.6511 | 0.1575 |
| 2007 | 153 | 0.1407 | 0.0305 | 85 | 0.0967 | 0.2232 |
| 2008 | 190 | 0.114 | 0.3734 | 91 | 0.2978 | 0.5806 |
| 2009 | 177 | 0.9668 | 0.1486 | 90 | 0.1886 | 0.1973 |
| 2010 | 12 | 0.9227 | 0.2357 | 6 | 1 | 0.4138 |
| 2011 | 229 | 0.5482 | 0.4589 | 14 | 0.8996 | 0.5606 |
| 2012 | 111 | 0.3796 | 0.8114 | 102 | 0.1864 | 0.8891 |

# Appendix F. Model outputs from Chapter 5

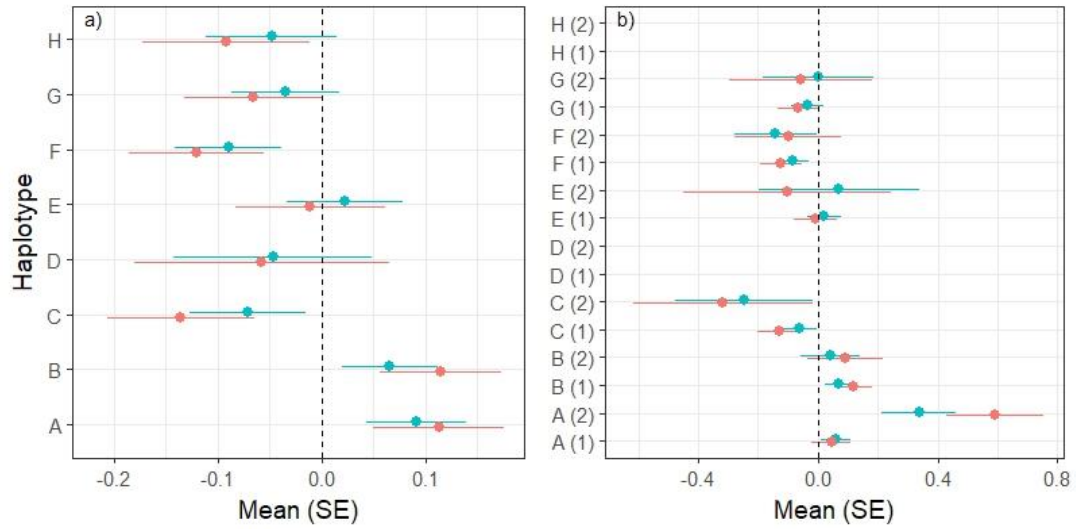



Figure F.0.1. Comparisons of estimates of haplotype effects on FEC from LMM of Log(Fec+50) (green) and GLMM of untransformed FEC with negative binomial errors (orange) in lambs as a) presence of haplotype and b) dosage of haplotype.

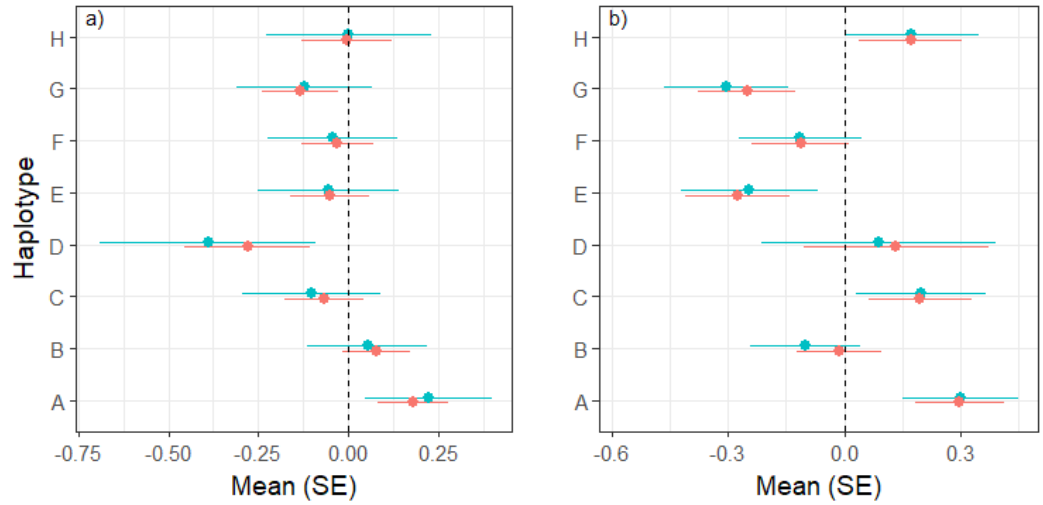



Figure F.0.2. Comparisons of estimates of haplotype effects on FEC from LMM of Log(Fec+50) (green) and GLMM of untransformed FEC with negative binomial errors (orange) in yearling females (a) and males (b) for presence of haplotype.

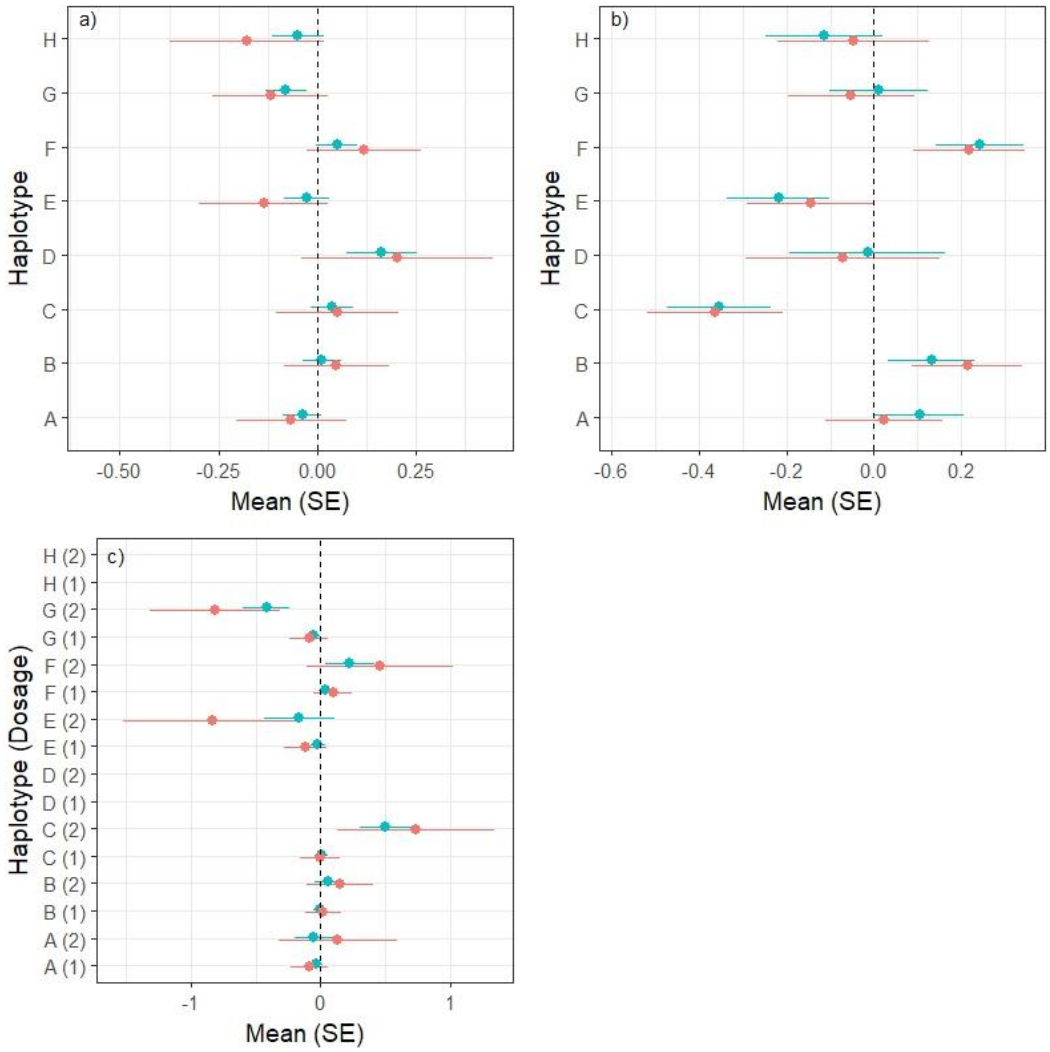



Figure F.0.3. Comparisons of estimates of haplotype effects on FEC from LMM of Log(Fec+50) (green) and GLMM of untransformed FEC with negative binomial errors (orange) in a) adult females (presence of haplotype), b) adult males (presence of haplotype) and c) adult females (dosage of haplotype).

Table F.1. Null models for lambs

| | | August weight | | Immunoglobulin | | | | | | FEC | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | IgA | | IgE | | IgG | | | |
| | | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value |
| GLMM | Intercept | 8.824 (0.334) | | 0.728 (0.030) | | 0.087 (0.007) | | 0.260 (0.014) | | 7.301 (0.193) | |
| | Sex: male | 1.398 (0.093) | < 0.0001 | -0.053 (0.023) | 0.0245 | 0.003 (0.006) | 0.6132 | -0.030 (0.009) | 0.0011 | 0.474 (0.062) | <0.0001 |
| | Litter size: twin | -3.199 (0.131) | < 0.0001 | 0.083 (0.031) | 0.0070 | -0.002 (0.007) | 0.778 | -0.024 (0.011) | 0.038 | 0.191 (0.079) | 0.0160 |
| | Age (Days)* | 0.050 (0.008) | < 0.0001 | 0.009 (0.002) | < 0.0001 | 0.003 (0.0005) | < 0.0001 | 0.006 (0.0007) | < 0.0001 | | |
| | Fhat3 | -4.193 (1.618) | 0.0095 | -0.618 (0.411) | 0.1343 | -0.145 (0.103) | 0.1635 | -0.298 (0.158) | 0.060 | -0.384 (0.993) | 0.6547 |
| | Maternal age | 1.744 (0.082) | | | | | | | | | |
| | Maternal age$^2$ | -0.137 (0.007) | < 0.0001 | | | | | | | | |
| | August weight* | | | | | | | | | -0.094 (0.013) | <0.0001 |
| | Maternal ID | 0.998 (0.999) | | 0.04 (0.201) | | 0.001 (0.033) | | 0.003 (0.053) | | 0.0000001 (0.0003) | |
| | Birth year | 1.642 (1.282) | | 0.012 (0.109) | | 0.0005 (0.022) | | 0.003 (0.052) | | 0.225 (0.474) | |
| | ELISA plate | | | 0.003 (0.052) | | 0.0003 (0.017) | | 0.001 (0.031) | | | |
| | Residual‡ | 2.646 (1.627) | | 0.183 (0.428) | | 0.012 (0.111) | | 0.028 (0.169) | | | |
| AM | Intercept | 8.873 (0.348) | | 0.720 (0.043) | | 0.079 (0.009) | | 0.254 (0.017) | | 7.013 (0.165) | |
| | Sex: male | 1.405 (0.092) | | -0.053 (0.022) | | 0.002 (0.006) | | -0.029 (0.009) | | 0.462 (0.046) | |
| | Litter size: twin | -3.214 (0.13) | | 0.078 (0.029) | | -0.003 (0.007) | | -0.025 (0.011) | | 0.167 (0.060) | |
| | Age (Days)* | 0.05 (0.008) | | 0.008 (0.002) | | 0.003 (0.001) | | 0.006 (0.001) | | | |
| | Fhat3 | -4.091 (1.614) | | -0.762 (0.392) | | -0.111 (0.102) | | -0.300 (0.155) | | 0.003 (0.782) | |
| | Maternal age | 1.742 (0.081) | | | | | | | | | |
| | Maternal age$^2$ | -0.137 (0.007) | | | | | | | | | |
| | August weight* | | | | | | | | | -0.083 (0.010) | |
| | $V_{Additive}$ | 0.635 (0.212) | | 0.095 (0.015) | | 0.002 (0.001) | | 0.008 (0.002) | | 0.152 (0.037) | |
| | $V_{Maternal}$ | 0.752 (0.139) | | 0.014 (0.006) | | 0.0005 (0.0003) | | 0.0003 (0.0008) | | 0.0000001 (0.0000) | |
| | $V_{BirthYear}$ | 1.713 (0.52) | | 0.012 (0.005) | | 0.0005 (0.0002) | | 0.003 (0.001) | | 0.149 (0.048) | |
| | $V_{Plate}$ | | | 0.003 (0.002) | | 0.0003 (0.0002) | | 0.001 (0.001) | | | |
| | $V_{Residual}$ | 2.237 (0.174) | | 0.113 (0.011) | | 0.011 (0.001) | | 0.022 (0.002) | | 0.551 (0.036) | |

* Mean centred

‡ not calculated within GLMM for FEC

Table F.2. Null models for yearlings

| | | August weight | | FEC | | | |
|---|---|---|---|---|---|---|---|
| | | | | Females | | Males | |
| | | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value |
| GLMM | Intercept | 17.318 (0.291) | | 7.466 (0.568) | | 7.873 (0.387) | |
| | Sex: male | 4.262 (0.188) | < 0.0001 | | | | |
| | Fhat3 | -7.684 (3.484) | 0.0276 | 1.688 (3.131) | 0.5839 | 2.616 (2.959) | 0.3711 |
| | Weight* | | | -0.114 (0.033) | 0.0007 | -0.081 (0.018) | < 0.0001 |
| | Year | 1.793 (1.339) | | 0.0000003 (0.0006) | | | |
| | Residual‡ | 7.085 (2.662) | | | | | |
| AM | Intercept | 17.307 (0.323) | | 6.928 (0.376) | | 6.981 (0.319) | |
| | Sex: male | 4.283 (0.184) | | | | | |
| | Fhat3 | -7.434 (3.429) | | 2.263 (1.695) | | 1.785 (2.128) | |
| | Weight* | | | -0.091 (0.021) | | -0.049 (0.014) | |
| | $V_{Additive}$ | 1.851 (0.549) | | 0.210 (0.093) | | 0.224 (0.139) | |
| | $V_{Year}$ | 1.829 (0.615) | | 0.059 (0.033) | | 0.058 (0.038) | |
| | $V_{Residual}$ | 5.306 (0.509) | | 0.565 (0.089) | | 0.531 (0.132) | |

* Mean centred

‡ not calculated within GLMM for FEC

Table F.3. Null models from GLMMs for adults

| | August weight | | Immunoglobulins | | | | | | FEC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ages included | 2 + | | 1 + | | | | | | 2 + | | | |
| | | | IgA | | IgE | | IgG | | Females | | Males | |
| | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value | Estimate (SE) | P-value |
| Intercept | 18.142 (0.269) | | 1.418 (0.031) | | 0.0493 (0.022) | | 0.738 (0.018) | | 5.395 (0.327) | | 6.181 (0.113) | |
| Sex: male | 9.773 (0.203) | < 0.0001 | -0.021 (0.036) | 0.5317 | -0.001 (0.024) | 0.9865 | -0.071 (0.012) | < 0.0001 | | | | |
| Age (Years) | 1.588 (0.055) | | 0.017 (0.003) | < 0.0001 | 0.098 (0.006) | | -0.042 (0.005) | | -0.377 (0.117) | | | |
| $Age^2$ | -0.107 (0.004) | < 0.0001 | | | -0.006 (0.001) | < 0.0001 | 0.002 (0.000) | < 0.0001 | 0.031 (0.010) | 0.0004 | | |
| Fhat3 | -8.392 (3.220) | 0.0092 | 0.313 (0.584) | 0.5879 | -0.342 (0.421) | 0.4163 | -0.157 (0.202) | 0.4387 | 1.577 (2.523) | 0.5376 | 0.103 (2.403) | 1.0000 |
| August weight* | | | | | | | | | -0.116 (0.026) | < 0.0001 | -0.028 (0.012) | 0.018 |
| ID | 6.050 (2.460) | | 0.237 (0.487) | | 0.120 (0.346) | | 0.024 (0.154) | | 0.00000012 (0.0003) | | 0.0000067 (0.0026) | |
| Birth year | 0.143 (0.378) | | 0.005 (0.072) | | | | 0.0003 (0.016) | | 0.00000367 (0.0019) | | | |
| Capture year | 1.203 (1.097) | | 0.002 (0.041) | | 0.003 (0.051) | | 0.002 (0.042) | | 0.000000113 (0.0003) | | 0.00836 (0.091) | |
| ELISA plate | | | 0.007 (0.083) | | 0.001 (0.035) | | 0.006 (0.075) | | | | | |
| Residual‡ | 2.178 (1.476) | | 0.058 (0.242) | | 0.038 (0.195) | | 0.015 (0.122) | | | | | |

* Mean centred

‡ not calculated within GLMM for FEC

Table F.4. Null models from AMs for adults

| | August weight | Immunoglobulins | | | FEC | |
|---|---|---|---|---|---|---|
| Ages included | 2 + | 1 + | | | 2 + | |
| | | IgA | IgE | IgG | Females | Males |
| | Estimate (SE) | Estimate (SE) | Estimate (SE) | Estimate (SE) | Estimate (SE) | Estimate (SE) |
| Intercept | 17.88 (0.318) | 1.405 (0.049) | 0.471 (0.034) | 0.718 (0.020) | 6.869 (0.203) | 6.351 (0.291) |
| Sex: male | 9.83 (0.196) | -0.025 (0.029) | 0.004 (0.022) | -0.071 (0.012) | | |
| Age (Years) | -7.434 (3.429) | 0.018 (0.003) | 0.098 (0.006) | -0.042 (0.005) | -0.214 (0.032) | |
| $Age^2$ | 1.602 (0.055) | | -0.006 (0.001) | 0.002 (0.0004) | 0.016 (0.002) | |
| Fhat3 | -0.107 (0.004) | 0.478 (0.500) | -0.119 (0.384) | -0.014 (0.197) | 0.293 (0.858) | -0.208 (1.866) |
| August weight* | -6.608 (3.075) | | | | -0.07 (0.009) | -0.019 (0.009) |
| $V_{Additive}$ | 2.529 (0.493) | 0.188 (0.021) | 0.077 (0.01) | 0.009 (0.002) | 0.058 (0.025) | 0.0000003 (0) |
| $V_{PermanentEnvironment}$ | 3.628 (0.415) | 0.051 (0.012) | 0.045 (0.007) | 0.015 (0.002) | 0.035 (0.025) | 0.174 (0.051) |
| $V_{BirthYear}$ | 0.259 (0.134) | 0.004 (0.003) | | 0.0001 (0.0003) | 0.005 (0.005) | 0.072 (0.032) |
| $V_{CaptureYear}$ | 1.139 (0.327) | 0.002 (0.001) | 0.003 (0.001) | 0.002 (0.001) | 0.018 (0.008) | |
| $V_{Plate}$ | | 0.007 (0.002) | 0.001 (0.001) | 0.006 (0.001) | | |
| $V_{Residual}$ | 2.314 (0.075) | 0.058 (0.002) | 0.038 (0.001) | 0.015 (0.0005) | 0.588 (0.023) | 0.527 (0.05) |

* Mean centred

Table F.5. Model output for August weight models of haplotype presence/absence in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| No interaction | | N (absent) | 1047 | 924 | 1185 | 1398 | 1196 | 1078 | 1117 | 1268 |
| | | N (present) | 451 | 574 | 313 | 100 | 302 | 420 | 381 | 230 |
| | LMM | Haplotype: 1 | 0.143 | 0.032 | 0.070 | -0.107 | -0.197 | -0.127 | 0.096 | 0.022 |
| | | | (0.106) | (0.102) | (0.124) | (0.200) | (0.123) | (0.109) | (0.112) | (0.135) |
| | | Log Likelihood | -3068.0 | -3068.8 | -3068.7 | -3068.8 | -3067.6 | -3068.2 | -3068.5 | -3068.9 |
| | | $\chi^2$ | 0.098 | 0.322 | 0.293 | 2.550 | 1.360 | 0.729 | 0.027 | 0.098 |
| | | P-value [a] | 0.7544 | 0.5702 | 0.5883 | 0.1103 | 0.2436 | 0.3932 | 0.8686 | 0.7544 |
| | AM | Haplotype: 1 | 0.156 | 0.102 | -0.008 | -0.176 | -0.204 | -0.148 | 0.09 | 0.038 |
| | | | (0.111) | (0.108) | (0.13) | (0.214) | (0.128) | (0.115) | (0.118) | (0.142) |
| | | Conditional Wald F-test [c] | 1.97 | 0.90 | 0.00 | 0.68 | 2.53 | 1.65 | 0.58 | 0.07 |
| | | P-value | 0.1608 | 0.3425 | 0.9502 | 0.4115 | 0.1117 | 0.1987 | 0.4480 | 0.7867 |
| With interaction | | N (absent) | F=539 M=508 | F=469 M=455 | F=602 M=583 | F=719 M=679 | F=614 M=582 | F=555 M=523 | F=577 M=540 | F=649 M=619 |
| | | N (present) | F=229 M=222 | F=299 M=275 | F=166 M=147 | F=49 M=51 | F=154 M=148 | F=213 M=207 | F=191 M=190 | F=119 M=111 |
| | LMM | Log Likelihood | -3066.9 | -3068.8 | -3067.3 | -3068.5 | -3067.2 | -3066.6 | -3068.4 | -3068.5 |
| | | $\chi^2$ | 2.05 | 0.14 | 2.78 | 0.44 | 0.82 | 3.25 | 0.30 | 0.71 |
| | | P-value [b] | 0.1518 | 0.7114 | 0.0955 | 0.5069 | 0.3660 | 0.0715 | 0.5846 | 0.4010 |
| | AM | Conditional Wald F-test [d] | 1.64 | 0.24 | 3.26 | 0.56 | 0.95 | 2.88 | 0.21 | 0.82 |
| | | P-value | 0.2002 | 0.6253 | 0.0713 | 0.4540 | 0.3308 | 0.0901 | 0.6492 | 0.3659 |

[a] compared to null model (log likelihood = -3068.9)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.6. Model output for August weight models of haplotype dosage in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1047 | 924 | 1185 | 1398 | 1196 | 1078 | 1117 | 1268 |
| | | N (1 copy) | 404 | 498 | 300 | 96 | 291 | 384 | 354 | 219 |
| | | N (2 copies) | 47 | 76 | 13 | 4 | 11 | 36 | 27 | 11 |
| | LMM | Haplotype: 1 | 0.117 | 0.050 | 0.123 | | -0.180 | -0.142 | 0.134 | -0.036 |
| | | | (0.109) | (0.105) | (0.125) | | (0.124) | (0.111) | (0.114) | (0.136) |
| | | Haplotype: 2 | 0.412 | -0.110 | -1.102 | | -0.704 | 0.073 | -0.474 | 1.334 |
| | | | (0.275) | (0.222) | (0.511) | | (0.543) | (0.312) | (0.356) | (0.561) |
| | | Log Likelihood | -3067.4 | -3068.6 | -3065.9 | | -3067.2 | -3068.0 | -3067.1 | -3066.0 |
| | | $\chi^2$ | 2.95 | 0.62 | 5.93 | | 3.48 | 1.83 | 3.60 | 5.84 |
| | | P-value [a] | 0.2286 | 0.7324 | 0.0515 | | 0.1758 | 0.4008 | 0.1654 | 0.0540 |
| | AM | Haplotype: 1 | 0.127 | 0.11 | 0.045 | | -0.19 | -0.16 | 0.124 | -0.015 |
| | | | (0.114) | (0.109) | (0.131) | | (0.129) | (0.116) | (0.12) | (0.143) |
| | | Haplotype: 2 | 0.495 | 0.01 | -1.435 | | -0.694 | 0.071 | -0.561 | 1.414 |
| | | | (0.28) | (0.234) | (0.515) | | (0.547) | (0.321) | (0.362) | (0.56) |
| | | Conditional Wald F-test [c] | 1.86 | 0.55 | 4.08 | | 1.69 | 1.09 | 2.10 | 3.26 |
| | | P-value [c] | 0.1566 | 0.5797 | 0.0171 | | 0.1849 | 0.3351 | 0.1229 | 0.0386 |
| With interaction | | N (absent) | F=539 M=508 | F=469 M=455 | F=602 M=583 | F=719 M=679 | F=614 M=582 | F=555 M=523 | F=577 M=540 | F=649 M=619 |
| | | N (1 copy) | F=203 M=201 | F=256 M=242 | F=160 M=140 | F=48 M=48 | F=148 M=143 | F=198 M=186 | F=177 M=177 | F=114 M=105 |
| | | N (2 copies) | F=26 M=21 | F=43 M=33 | F=6 M=7 | F=1 M=3 | F=6 M=5 | F=15 M=21 | F=14 M=13 | F=5 M=6 |
| | LMM | Log Likelihood | -3066.3 | -3068.0 | | | | -3066.1 | -3066.9 | |
| | | $\chi^2$ | 2.21 | 1.12 | | | | 3.72 | 0.35 | |
| | | P-value [b] | 0.3309 | 0.5724 | | | | 0.1554 | 0.8394 | |
| | AM | Conditional Wald F-test [d] | 0.93 | 0.88 | | | | 1.71 | 0.21 | |
| | | P-value [d] | 0.3958 | 0.4166 | | | | 0.1815 | 0.6492 | |

[a] compared to null model (log likelihood = -3068.9)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.7. Model output for August weight models of haplotype presence/absence in yearlings.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| No interaction | | N (absent) | 598 | 531 | 672 | 812 | 690 | 640 | 650 | 717 |
| | | N (present) | 268 | 335 | 194 | 54 | 176 | 226 | 216 | 149 |
| | LMM | Haplotype: 1 | 0.095 | 0.184 | -0.129 | 0.405 | -0.756 | -0.231 | 0.712 | -0.221 |
| | | | (0.198) | (0.187) | (0.219) | (0.381) | (0.226) | (0.210) | (0.213) | (0.246) |
| | | Log Likelihood | -2104.5 | -2104.1 | -2104.0 | -2104.0 | -2099.0 | -2104.0 | -2099.0 | -2104.2 |
| | | $\chi^2$ | 0.23 | 0.96 | 0.34 | 1.13 | 11.12 | 1.21 | 11.09 | 0.81 |
| | | P-value [a] | 0.6320 | 0.3261 | 0.5594 | 0.2883 | 0.0009 | 0.2710 | 0.0009 | 0.3685 |
| | AM | Haplotype: 1 | 0.101 | 0.245 | -0.085 | 0.243 | -0.738 | -0.279 | 0.688 | -0.091 |
| | | | (0.207) | (0.203) | (0.241) | (0.417) | (0.239) | (0.224) | (0.231) | (0.262) |
| | | Conditional Wald F-test [c] | 0.24 | 1.45 | 0.12 | 0.34 | 9.54 | 1.57 | 8.87 | 0.12 |
| | | P-value | 0.6275 | 0.2284 | 0.7259 | 0.5604 | 0.0021 | 0.2110 | 0.0030 | 0.7281 |
| With interaction | | N (absent) | F=352 M=246 | F=321 M=210 | F=396 M=276 | F=476 M=336 | F=398 M=292 | F=375 M=265 | F=378 M=272 | F=435 M=282 |
| | | N (present) | F=159 M=109 | F=190 M=145 | F=115 M=79 | F=35 M=19 | F=113 M=63 | F=136 M=90 | F=133 M=83 | F=76 M=73 |
| | LMM | Log Likelihood | -2104.5 | 2104.0 | -2104.4 | -2104.0 | -2098.9 | -2102.6 | -2098.0 | -2102.6 |
| | | $\chi^2$ | 0.04 | 0.21 | 0.03 | 0.00 | 0.18 | 2.83 | 2.10 | 1.53 |
| | | P-value [b] | 0.8511 | 0.6451 | 0.8685 | 0.9518 | 0.6741 | 0.0927 | 0.1476 | 0.2158 |
| | AM | Conditional Wald F-test [d] | 0.00 | 0.05 | 0.18 | 0.00 | 0.02 | 3.76 | 1.52 | 1.56 |
| | | P-value | 0.9926 | 0.8189 | 0.6741 | 0.9597 | 0.8982 | 0.0529 | 0.2175 | 0.2126 |

[a] compared to null model (log likelihood = -2104.6)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Appendix F

Table F.8. Model output for August weight models of haplotype dosage in yearlings.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 598 | 531 | 672 | 812 | 690 | 640 | 650 | 717 |
| | | N (1 copy) | 248 | 298 | 183 | 54 | 168 | 209 | 202 | 142 |
| | | N (2 copies) | 20 | 37 | 11 | 0 | 8 | 17 | 14 | 7 |
| | LMM | Haplotype: 1 | 0.164 | 0.183 | -0.144 | | | -0.262 | 0.712 | |
| | | | (0.203) | (0.194) | (0.224) | | | (0.217) | (0.219) | |
| | | Haplotype: 2 | -0.703 | 0.195 | 0.139 | | | 0.145 | 0.708 | |
| | | | (0.585) | (0.458) | (0.827) | | | (0.664) | (0.785) | |
| | | Log Likelihood | -2103.4 | 2104.1 | -2104.4 | | | -2103.8 | -2099.0 | |
| | | $\chi^2$ | 2.34 | 0.97 | 0.46 | | | 1.57 | 11.09 | |
| | | P-value [a] | 0.3098 | 0.6172 | 0.7965 | | | 0.4570 | 0.0039 | |
| | AM | Haplotype: 1 | 0.177 | 0.244 | -0.103 | | | -0.296 | 0.695 | |
| | | | (0.211) | (0.206) | (0.244) | | | (0.227) | (0.234) | |
| | | Haplotype: 2 | -0.981 | 0.254 | 0.31 | | | 0.016 | 0.54 | |
| | | | (0.586) | (0.471) | (0.829) | | | (0.67) | (0.788) | |
| | | Conditional Wald F-test [c] | 2.08 | 0.73 | 0.19 | | | 0.89 | 4.44 | |
| | | P-value [c] | 0.1258 | 0.4843 | 0.8311 | | | 0.4092 | 0.0121 | |
| With interaction | | N (absent) | F=352 M=246 | F=321 M=210 | F=396 M=276 | F=476 M=336 | F=398 M=292 | F=375 M=265 | F=378 M=272 | F=435 M=282 |
| | | N (1 copy) | F=149 M=99 | F=167 M=131 | F=108 M=75 | F=35 M=19 | F=107 M=61 | F=128 M=81 | F=124 M=78 | F=74 M=68 |
| | | N (2 copies) | F=10 M=10 | F=23 M=14 | F=7 M=4 | F=0 M=0 | F=6 M=2 | F=8 M=9 | F=9 M=5 | F=2 M=5 |
| LMM | | Haplotype: 1 | 0.116 | 0.081 | | | | | | |
| | | | (0.259) | (0.253) | | | | | | |
| | | Haplotype: 2 | 0.199 | 0.356 | | | | | | |
| | | | (0.862) | (0.582) | | | | | | |
| | | Haplotype: 1 * Sex:Male | 0.126 | 0.248 | | | | | | |
| | | | (0.418) | (0.398) | | | | | | |
| | | Haplotype:2 * Sex:Male | -1.658 | -0.423 | | | | | | |
| | | | (1.171) | (0.942) | | | | | | |
| | | Log Likelihood | -2102.3 | -2103.8 | | | | | | |
| | | $\chi^2$ | 2.22 | 0.70 | | | | | | |
| | | P-value [b] | 0.3297 | 0.7060 | | | | | | |
| AM | | Conditional Wald F-test [d] | 1.57 | 0.39 | | | | | | |
| | | P-value [d] | 0.2078 | 0.6795 | | | | | | |

[a] compared to null model (log likelihood = -2104.6)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.9. Model output for August weight models of haplotype presence/absence in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 628 | 579 | 698 | 871 | 747 | 682 | 705 | 795 |
| | | N (present) | 300 | 349 | 230 | 57 | 181 | 246 | 223 | 133 |
| | LMM | Haplotype: 1 | -0.158 | 0.094 | -0.636 | -0.010 | -0.172 | 0.219 | 0.586 | -0.107 |
| | | | (0.187) | (0.181) | (0.203) | (0.379) | (0.219) | (0.200) | (0.206) | (0.249) |
| | | Log Likelihood | -6437.0 | 6437.3 | -6432.5 | -6437.4 | -6437.1 | -6436.8 | -6433.4 | -6437.3 |
| | | $\chi^2$ | 0.71 | 0.28 | 9.76 | 0.00 | 0.63 | 1.21 | 8.09 | 0.19 |
| | | P-value [a] | 0.3983 | 0.6000 | 0.0018 | 0.9731 | 0.4289 | 0.2720 | 0.0044 | 0.6670 |
| | AM | Haplotype: 1 | -0.164 | 0.047 | -0.618 | -0.11 | 0.032 | 0.192 | 0.359 | -0.012 |
| | | | (0.198) | (0.199) | (0.225) | (0.427) | (0.236) | (0.215) | (0.226) | (0.265) |
| | | Conditional Wald F-test [c] | 0.68 | 0.05 | 7.53 | 0.07 | 0.02 | 0.80 | 2.53 | 0.00 |
| | | P-value | 0.4098 | 0.8150 | 0.0062 | 0.7961 | 0.8939 | 0.3715 | 0.1119 | 0.9649 |
| With interaction | | N (absent) | F=456 M=172 | F=422 M=157 | F=504 M=194 | F=634 M=237 | F=536 M=211 | F=503 M=179 | F=509 M=196 | F=580 M=215 |
| | | N (present) | F=218 M=82 | F=252 M=97 | F=170 M=60 | F=40 M=17 | F=138 M=43 | F=171 M=75 | F=165 M=58 | F=94 M=39 |
| | LMM | Log Likelihood | -6437.0 | -6437.3 | 6432.2 | -6436.4 | -6437.1 | -6436.4 | -6439.0 | -6437.3 |
| | | $\chi^2$ | 0.17 | 0.03 | 0.59 | 2.08 | 0.03 | 0.84 | 3.45 | 0.05 |
| | | P-value [b] | 0.6793 | 0.8589 | 0.4412 | 0.1493 | 0.8553 | 0.3594 | 0.0634 | 0.8231 |
| | AM | Conditional Wald F-test [d] | 0.01 | 0.06 | 2.07 | 2.37 | 0.00 | 1.66 | 3.05 | 0.00 |
| | | P-value | 0.9118 | 0.8003 | 0.1501 | 0.1238 | 0.9820 | 0.1978 | 0.0809 | 0.9746 |

[a] compared to null model (log likelihood = -6437.4)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.10. Model output for August weight models of haplotype dosage in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 628 | 579 | 698 | 871 | 747 | 682 | 705 | 795 |
| | | N (1 copy) | 277 | 300 | 214 | 56 | 173 | 227 | 210 | 125 |
| | | N (2 copies) | 23 | 49 | 16 | 1 | 8 | 19 | 13 | 8 |
| | LMM | Haplotype: 1 | -0.176 | 0.023 | -0.627 | | -0.158 | 0.213 | 0.562 | -0.077 |
| | | | (0.193) | (0.190) | (0.208) | | (0.223) | (0.206) | (0.211) | (0.254) |
| | | Haplotype: 2 | 0.056 | 0.525 | -0.752 | | -0.488 | 0.290 | 0.952 | -0.697 |
| | | | (0.557) | (0.394) | (0.673) | | (0.985) | (0.622) | (0.762) | (1.066) |
| | | Log Likelihood | -6437.0 | -6436.5 | -6432.5 | | -6437.0 | -6436.8 | -6433.2 | -6437.2 |
| | | $\chi^2$ | 0.88 | 1.79 | 9.80 | | 0.74 | 1.22 | 8.34 | 0.51 |
| | | P-value [a] | 0.6444 | 0.4084 | 0.0075 | | 0.6922 | 0.5431 | 0.0155 | 0.7738 |
| | AM | Haplotype: 1 | -0.177 | 0.007 | -0.632 | | 0.028 | 0.180 | 0.366 | 0.013 |
| | | | (0.201) | (0.201) | (0.227) | | (0.238) | (0.217) | (0.227) | (0.269) |
| | | Haplotype: 2 | 0.071 | 0.509 | -0.310 | | 0.177 | 0.465 | 0.147 | -0.550 |
| | | | (0.564) | (0.412) | (0.670) | | (0.948) | (0.638) | (0.759) | (1.044) |
| | | Conditional Wald F-test [c] | 0.44 | 0.85 | 3.87 | | 0.02 | 0.50 | 1.31 | 0.14 |
| | | P-value [c] | 0.6443 | 0.4293 | 0.0212 | | 0.9780 | 0.6055 | 0.2697 | 0.8672 |
| With interaction | | N (absent) | F=456 M=172 | F=422 M=157 | F=504 M=194 | F=634 M=237 | F=536 M=211 | F=503 M=179 | F=509 M=196 | F=580 M=215 |
| | | N (1 copy) | F=201 M=76 | F=212 M=88 | F=160 M=54 | F=39 M=17 | F=133 M=40 | F=160 M=67 | F=154 M=56 | F=89 M=36 |
| | | N (2 copies) | F=17 M=6 | F=40 M=9 | F=10 M=6 | F=1 M=0 | F=5 M=3 | F=11 M=8 | F=11 M=2 | F=5 M=3 |
| | LMM | Log Likelihood | -6435.5 | -6436.4 | -6432.1 | | | -6435.4 | -6431.3 | |
| | | $\chi^2$ | 2.90 | 0.13 | 0.73 | | | 2.86 | 3.94 | |
| | | P-value [b] | 0.2346 | 0.9388 | 0.6951 | | | 0.2399 | 0.1393 | |
| | AM | Conditional Wald F-test [d] | 2.37 | 0.19 | 1.27 | | | 1.74 | 1.67 | |
| | | P-value [d] | 0.0945 | 0.8245 | 0.2804 | | | 0.1757 | 0.1899 | |

[a] compared to null model (log likelihood = -6437.4)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.11. Model output for anti-*T. circ* IgA models of haplotype presence/absence in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| No interaction | | N (absent) | 1084 | 958 | 1256 | 1469 | 1260 | 1143 | 1177 | 1334 |
| | | N (present) | 489 | 615 | 317 | 104 | 313 | 430 | 396 | 239 |
| | LMM | Haplotype: 1 | 0.007 | -0.006 | 0.092 | 0.021 | 0.059 | -0.066 | -0.026 | -0.007 |
| | | | (0.026) | (0.025) | (0.031) | (0.050) | (0.031) | (0.027) | (0.028) | (0.034) |
| | | Log Likelihood | -1021.3 | -1021.3 | -1016.9 | -1021.3 | -1019.5 | -1018.4 | -1020.9 | -1021.3 |
| | | $\chi^2$ | 0.08 | 0.06 | 8.95 | 0.18 | 3.71 | 5.82 | 0.85 | 0.04 |
| | | P-value [a] | 0.7817 | 0.8124 | 0.0028 | 0.6726 | 0.0542 | 0.0158 | 0.3560 | 0.8366 |
| | AM | Haplotype: 1 | 0.046 | -0.016 | 0.066 | 0.017 | 0.086 | -0.033 | -0.074 | -0.042 |
| | | | (0.028) | (0.027) | (0.033) | (0.055) | (0.032) | (0.029) | (0.030) | (0.036) |
| | | Conditional Wald F-test [c] | 2.79 | 0.34 | 4.09 | 0.09 | 7.15 | 1.24 | 6.19 | 1.34 |
| | | P-value | 0.0953 | 0.5605 | 0.0433 | 0.7602 | 0.0076 | 0.2664 | 0.0130 | 0.2469 |
| With interaction | | N (absent) | F=551 M=533 | F=490 M=468 | F=638 M=618 | F=751 M=718 | F=647 M=613 | F=587 M=556 | F=609 M=568 | F=684 M=650 |
| | | N (present) | F=254 M=235 | F=315 M=300 | F=167 M=150 | F=54 M=50 | F=158 M=155 | F=218 M=212 | F=196 M=200 | F=121 M=118 |
| | LMM | Log Likelihood | -1020.3 | -1021.3 | -1016.4 | -1021.2 | -1018.8 | -1017.6 | -1019.8 | -1019.2 |
| | | $\chi^2$ | 2.00 | 0.02 | 0.98 | 0.02 | 1.33 | 1.64 | 2.25 | 4.18 |
| | | P-value [b] | 0.1577 | 0.8828 | 0.3214 | 0.8887 | 0.2489 | 0.2001 | 0.1334 | 0.0408 |
| | AM | Conditional Wald F-test [d] | 2.79 | 0.34 | 4.09 | 0.09 | 7.15 | 1.24 | 6.19 | 1.34 |
| | | P-value | 0.0953 | 0.5605 | 0.0433 | 0.7602 | 0.0076 | 0.2664 | 0.0130 | 0.2469 |

[a] compared to null model (log likelihood = -1021.4)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.12. Model output for anti-*T. circ* IgA models of haplotype dosage in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1084 | 958 | 1256 | 1469 | 1260 | 1143 | 1177 | 1334 |
| | | N (1 copy) | 437 | 530 | 303 | 100 | 299 | 394 | 368 | 229 |
| | | N (2 copies) | 52 | 85 | 14 | 4 | 14 | 36 | 28 | 10 |
| | LMM | Haplotype: 1 | 0.020 | -0.005 | 0.087 | | 0.059 | -0.061 | -0.025 | -0.007 |
| | | | (0.027) | (0.026) | (0.031) | | (0.031) | (0.028) | (0.029) | (0.034) |
| | | Haplotype: 2 | -0.108 | -0.018 | 0.208 | | 0.080 | -0.137 | -0.047 | -0.006 |
| | | | (0.067) | (0.055) | (0.126) | | (0.125) | (0.079) | (0.094) | (0.149) |
| | | Log Likelihood | -1019.5 | -1021.3 | -1016.4 | | -1019.5 | -1018.0 | -1020.9 | -1021.3 |
| | | $\chi^2$ | 3.619 | 0.113 | 9.855 | | 3.738 | 6.708 | 0.906 | 0.043 |
| | | P-value [a] | 0.1637 | 0.9453 | 0.0072 | | 0.1543 | 0.0349 | 0.6358 | 0.9789 |
| | AM | Haplotype: 1 | 0.055 | -0.014 | 0.064 | | 0.086 | -0.032 | -0.074 | -0.042 |
| | | | (0.028) | (0.027) | (0.033) | | (0.032) | (0.03) | (0.03) | (0.036) |
| | | Haplotype: 2 | -0.06 | -0.043 | 0.139 | | 0.118 | -0.043 | -0.088 | -0.039 |
| | | | (0.067) | (0.058) | (0.123) | | (0.122) | (0.08) | (0.093) | (0.143) |
| | | Conditional Wald F-test [c] | 2.90 | 0.31 | 2.23 | | 3.61 | 0.63 | 3.10 | 0.67 |
| | | P-value [c] | 0.0552 | 0.7302 | 0.1078 | | 0.0273 | 0.5339 | 0.0453 | 0.5117 |
| With interaction | | N (absent) | F=551 M=533 | F=490 M=468 | F=638 M=618 | F=751 M=718 | F=647 M=613 | F=587 M=556 | F=609 M=568 | F=684 M=650 |
| | | N (1 copy) | F=224 M=213 | F=266 M=264 | F=161 M=142 | F=53 M=47 | F=150 M=149 | F=203 M=191 | F=183 M=185 | F=116 M=113 |
| | | N (2 copies) | F=30 M=22 | F=49 M=36 | F=6 M=8 | F=1 M=3 | F=8 M=6 | F=15 M=21 | F=13 M=15 | F=5 M=5 |
| | LMM | Log Likelihood | 1131.0 | 1128.6 | | | | 1131.7 | 1128.7 | |
| | | $\chi^2$ | 1.82 | 0.27 | | | | 1.41 | 0.31 | |
| | | P-value [b] | 0.178 | 0.604 | | | | 0.236 | 0.580 | |
| | AM | Conditional Wald F-test [d] | 1.18 | 0.36 | | | | 1.23 | 2.39 | |
| | | P-value [d] | 0.3082 | 0.6992 | | | | 0.2917 | 0.0918 | |

[a] compared to null model (log likelihood = -1021.4)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.13. Model output for anti-*T. circ* IgA models of haplotype presence/absence in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 708 | 876 | 1078 | 917 | 845 | 871 | 956 |
| | | N (present) | 371 | 434 | 266 | 64 | 225 | 297 | 271 | 186 |
| | LMM | Haplotype: 1 | -0.105 | 0.058 | 0.127 | 0.164 | -0.031 | -0.104 | 0.038 | -0.003 |
| | | | (0.033) | (0.032) | (0.037) | (0.067) | (0.039) | (0.036) | (0.037) | (0.043) |
| | | Log Likelihood | -1316.1 | -1319.4 | -1315.1 | -1318.0 | -1320.7 | -1316.7 | -1320.5 | -1321.0 |
| | | $\chi^2$ | 9.88 | 3.28 | 11.73 | 5.99 | 0.62 | 8.54 | 1.07 | 0.00 |
| | | P-value [a] | 0.0017 | 0.0701 | 0.0006 | 0.0144 | 0.4297 | 0.0035 | 0.3008 | 0.9502 |
| | AM | Haplotype: 1 | -0.034 | 0.065 | 0.063 | 0.069 | -0.009 | -0.069 | -0.035 | -0.006 |
| | | | (0.035) | (0.035) | (0.041) | (0.076) | (0.042) | (0.038) | (0.04) | (0.046) |
| | | Conditional Wald F-test [c] | 0.94 | 3.51 | 2.38 | 0.81 | 0.05 | 3.30 | 0.79 | 0.02 |
| | | P-value | 0.3337 | 0.0614 | 0.1235 | 0.3679 | 0.8206 | 0.0694 | 0.3745 | 0.8966 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (present) | | | | | | | | |
| | LMM | Log Likelihood | -1315.7 | -1319.4 | 1314.5 | -1316.9 | -1320.3 | -1316.7 | -1320.1 | -1321.0 |
| | | $\chi^2$ | 0.81 | 0.00 | 1.33 | 2.27 | 0.77 | 0.03 | 0.76 | 0.03 |
| | | P-value [b] | 0.3693 | 0.9949 | 0.2480 | 0.1317 | 0.3799 | 0.8681 | 0.3837 | 0.8594 |
| | AM | Conditional Wald F-test [d] | 1.26 | 0.01 | 1.06 | 3.69 | 2.33 | 0.14 | 0.32 | 0.11 |
| | | P-value | 0.2625 | 0.9227 | 0.3047 | 0.0552 | 0.1274 | 0.7046 | 0.5720 | 0.7369 |

[a] compared to null model (log likelihood = -1321.0)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.14. Model output for anti-*T. circ* IgA models of haplotype dosage in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 708 | 876 | 1078 | 917 | 845 | 871 | 956 |
| | | N (1 copy) | 337 | 379 | 249 | 63 | 214 | 275 | 253 | 174 |
| | | N (2 copies) | 34 | 55 | 17 | 1 | 11 | 22 | 18 | 12 |
| | LMM | Haplotype: 1 | -0.092 | 0.046 | 0.128 | | -0.030 | -0.096 | 0.048 | -0.006 |
| | | | (0.034) | (0.034) | (0.038) | | (0.040) | (0.037) | (0.038) | (0.044) |
| | | Haplotype: 2 | -0.232 | 0.141 | 0.114 | | -0.043 | -0.202 | -0.111 | 0.045 |
| | | | (0.092) | (0.074) | (0.129) | | (0.160) | (0.113) | (0.132) | (0.155) |
| | | Log Likelihood | -1315.0 | -1318.6 | -1315.1 | | -1320.7 | -1316.3 | -1319.8 | -1321.0 |
| | | $\chi^2$ | 12.09 | 4.86 | 11.74 | | 0.63 | 9.38 | 2.45 | 0.11 |
| | | P-value [a] | 0.0024 | 0.0879 | 0.0028 | | 0.7301 | 0.0092 | 0.2941 | 0.9470 |
| | AM | Haplotype: 1 | -0.031 | 0.065 | 0.062 | | -0.01 | -0.069 | -0.034 | -0.008 |
| | | | (0.035) | (0.035) | (0.041) | | (0.042) | (0.038) | (0.04) | (0.046) |
| | | Haplotype: 2 | -0.104 | 0.119 | 0.116 | | 0.001 | -0.059 | -0.118 | 0.058 |
| | | | (0.089) | (0.073) | (0.119) | | (0.14) | (0.107) | (0.121) | (0.145) |
| | | Conditional Wald F-test [c] | 0.84 | 2.10 | 1.30 | | 0.03 | 1.65 | 0.66 | 0.12 |
| | | P-value [c] | 0.4337 | 0.1225 | 0.2741 | | 0.9713 | 0.1918 | 0.5192 | 0.8907 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (1 copy) | | | | | | | | |
| | | N (2 copies) | | | | | | | | |
| | LMM | Log Likelihood | -1313.9 | -1318.2 | | | | -1316.3 | | |
| | | $\chi^2$ | 2.06 | 0.71 | | | | 0.07 | | |
| | | P-value [b] | 0.3569 | 0.7015 | | | | 0.9645 | | |
| | AM | Conditional Wald F-test [d] | 1.42 | 0.18 | | | | 0.07 | | |
| | | P-value [d] | 0.2427 | 0.8379 | | | | 0.9328 | | |

[a] compared to null model (log likelihood = -1321.0)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.15. Model output for anti-*T. circ* IgE models of haplotype presence/absence in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1088 | 962 | 1261 | 1474 | 1264 | 1144 | 1181 | 1338 |
| | | N (present) | 490 | 616 | 317 | 104 | 314 | 434 | 397 | 240 |
| | LMM | Haplotype: 1 | -0.013 | 0.006 | -0.005 | 0.026 | 0.002 | 0.016 | -0.007 | -0.005 |
| | | | (0.006) | (0.006) | (0.008) | (0.012) | (0.008) | (0.007) | (0.007) | (0.008) |
| | | Log Likelihood | 1130.1 | 1128.5 | 1128.2 | 1130.3 | 1128.0 | 1131.0 | 1128.6 | 1128.2 |
| | | $\chi^2$ | 4.16 | 0.91 | 0.40 | 4.60 | 0.07 | 5.93 | 1.12 | 0.36 |
| | | P-value [a] | 0.0414 | 0.3399 | 0.5293 | 0.0319 | 0.7964 | 0.0149 | 0.2902 | 0.5491 |
| | AM | Haplotype: 1 | -0.011 | 0.011 | -0.003 | 0.025 | -0.001 | 0.016 | -0.013 | -0.005 |
| | | | (0.007) | (0.007) | (0.008) | (0.013) | (0.008) | (0.007) | (0.007) | (0.009) |
| | | Conditional Wald F-test [c] | 2.50 | 2.58 | 0.16 | 3.59 | 0.02 | 5.07 | 3.03 | 0.36 |
| | | P-value | 0.1142 | 0.1085 | 0.6863 | 0.0583 | 0.8909 | 0.0245 | 0.0819 | 0.5495 |
| With interaction | | N (absent) | F=552 M=536 | F=490 M=472 | F=639 M=622 | F=752 M=722 | F=647 M=617 | F=588 M=556 | F=610 M=571 | F=685 M=653 |
| | | N (present) | F=254 M=536 | F=316 M=472 | F=167 M=622 | F=54 M=722 | F=159 M=617 | F=218 M=556 | F=196 M=571 | F=121 M=653 |
| | LMM | Log Likelihood | 1131.0 | 1128.6 | 1128.3 | 1130.4 | 1128.0 | 1131.7 | 1128.7 | 1128.5 |
| | | $\chi^2$ | 1.82 | 0.27 | 0.19 | 0.26 | 0.01 | 1.41 | 0.31 | 0.59 |
| | | P-value [b] | 0.1777 | 0.6039 | 0.6643 | 0.6134 | 0.9310 | 0.2355 | 0.5797 | 0.4425 |
| | AM | Conditional Wald F-test [d] | 1.71 | 0.11 | 0.09 | 0.30 | 0.01 | 1.36 | 0.56 | 0.61 |
| | | P-value | 0.1915 | 0.7350 | 0.7619 | 0.5830 | 0.9143 | 0.2445 | 0.4527 | 0.4363 |

[a] compared to null model (log likelihood = 1128.0)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.16. Model output for anti-*T. circ* IgE models of haplotype dosage in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1088 | 962 | 1261 | 1474 | 1264 | 1144 | 1181 | 1338 |
| | | N (1 copy) | 438 | 531 | 303 | 100 | 300 | 397 | 369 | 230 |
| | | N (2 copies) | 52 | 85 | 14 | 4 | 14 | 37 | 28 | 10 |
| | LMM | Haplotype: 1 | -0.011 | 0.008 | -0.005 | | 0.0005 | 0.016 | -0.006 | -0.005 |
| | | | (0.007) | (0.006) | (0.008) | | (0.008) | (0.007) | (0.007) | (0.008) |
| | | Haplotype: 2 | -0.033 | -0.007 | -0.004 | | 0.037 | 0.022 | -0.028 | 0.006 |
| | | | (0.017) | (0.014) | (0.031) | | (0.031) | (0.020) | (0.023) | (0.037) |
| | | Log Likelihood | 1130.9 | 1129.0 | 1128.2 | | 1128.7 | 1131.0 | 1129.0 | 1128.2 |
| | | $\chi^2$ | 5.84 | 2.07 | 0.40 | | 1.41 | 6.03 | 1.99 | 0.46 |
| | | P-value [a] | 0.0539 | 0.3557 | 0.8198 | | 0.4948 | 0.0491 | 0.3697 | 0.7944 |
| | AM | Haplotype: 1 | -0.009 | 0.012 | -0.003 | | -0.002 | 0.016 | -0.012 | -0.006 |
| | | | (0.007) | (0.007) | (0.008) | | (0.008) | (0.007) | (0.007) | (0.009) |
| | | Haplotype: 2 | -0.029 | -0.002 | -0.006 | | 0.031 | 0.019 | -0.032 | 0.008 |
| | | | (0.017) | (0.014) | (0.032) | | (0.031) | (0.02) | (0.024) | (0.037) |
| | | Conditional Wald F-test [c] | 1.96 | 1.82 | 0.09 | | 0.58 | 2.55 | 1.89 | 0.25 |
| | | P-value [c] | 0.1420 | 0.1632 | 0.9177 | | 0.5620 | 0.0788 | 0.1509 | 0.7788 |
| With interaction | | N (absent) | F=552 M=536 | F=490 M=472 | F=639 M=622 | F=752 M=722 | F=647 M=617 | F=588 M=556 | F=610 M=571 | F=685 M=653 |
| | | N (1 copy) | F=224 M=214 | F=267 M=264 | F=161 M=142 | F=53 M=47 | F=151 M=149 | F=203 M=194 | F=183 M=186 | F=116 M=114 |
| | | N (2 copies) | F=30 M=22 | F=49 M=36 | F=6 M=8 | F=1 M=3 | F=8 M=6 | F=15 M=22 | F=13 M=15 | F=5 M=5 |
| | LMM | Log Likelihood | 1132.8 | 1129.2 | | | | 1131.8 | 1129.2 | |
| | | $\chi^2$ | 3.83 | 0.28 | | | | 1.47 | 0.47 | |
| | | P-value [b] | 0.1473 | 0.8697 | | | | 0.4796 | 0.7918 | |
| | AM | Conditional Wald F-test [d] | 1.99 | 0.06 | | | | 0.74 | 0.51 | |
| | | P-value [d] | 0.1366 | 0.9378 | | | | 0.4770 | 0.5979 | |

[a] compared to null model (log likelihood = 1128.0)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.17. Model output for anti-*T. circ* IgE models of haplotype presence/absence in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 708 | 876 | 1078 | 917 | 845 | 871 | 956 |
| | | N (present) | 371 | 434 | 266 | 64 | 225 | 297 | 271 | 186 |
| | LMM | Haplotype: 1 | -0.091 | 0.086 | 0.048 | 0.105 | 0.025 | 0.006 | -0.081 | -0.035 |
| | | | (0.024) | (0.023) | (0.027) | (0.048) | (0.028) | (0.026) | (0.027) | (0.031) |
| | | Log Likelihood | -521.6 | -521.8 | -527.1 | -526.3 | -528.3 | -528.7 | -524.1 | -528.1 |
| | | $\chi^2$ | 14.19 | 13.75 | 3.27 | 4.77 | 0.79 | 0.06 | 9.24 | 1.27 |
| | | P-value [a] | 0.0002 | 0.0002 | 0.0706 | 0.0289 | 0.3728 | 0.8125 | 0.0024 | 0.2591 |
| | AM | Haplotype: 1 | -0.05 | 0.079 | 0.049 | 0.123 | 0.01 | -0.012 | -0.069 | -0.064 |
| | | | (0.026) | (0.026) | (0.03) | (0.056) | (0.031) | (0.028) | (0.029) | (0.034) |
| | | Conditional Wald F-test [c] | 3.71 | 9.35 | 2.72 | 4.93 | 0.11 | 0.20 | 5.52 | 3.67 |
| | | P-value | 0.0542 | 0.0023 | 0.0995 | 0.0267 | 0.7403 | 0.6583 | 0.0190 | 0.0557 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (present) | | | | | | | | |
| | LMM | Log Likelihood | -521.5 | -521.6 | -524.5 | -525.4 | -528.3 | -528.0 | -523.9 | -528.0 |
| | | $\chi^2$ | 0.18 | 0.51 | 5.06 | 1.89 | 0.08 | 1.27 | 0.45 | 0.05 |
| | | P-value [b] | 0.6701 | 0.4749 | 0.0244 | 0.1691 | 0.7816 | 0.2589 | 0.5028 | 0.8147 |
| | AM | Conditional Wald F-test [d] | 0.13 | 0.30 | 2.90 | 0.55 | 0.01 | 1.19 | 0.33 | 0.07 |
| | | P-value | 0.7138 | 0.5825 | 0.0889 | 0.4583 | 0.9244 | 0.2756 | 0.5663 | 0.7902 |

[a] compared to null model (log likelihood = -528.71)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.18. Model output for anti-*T. circ* IgE models of haplotype dosage in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 708 | 876 | 1078 | 917 | 845 | 871 | 956 |
| | | N (1 copy) | 337 | 379 | 249 | 63 | 214 | 275 | 253 | 174 |
| | | N (2 copies) | 34 | 55 | 17 | 1 | 11 | 22 | 18 | 12 |
| | LMM | Haplotype: 1 | -0.088 | 0.077 | 0.044 | | 0.028 | 0.006 | -0.065 | -0.025 |
| | | | (0.025) | (0.024) | (0.027) | | (0.029) | (0.027) | (0.027) | (0.032) |
| | | Haplotype: 2 | -0.119 | 0.149 | 0.119 | | -0.022 | 0.007 | -0.323 | -0.168 |
| | | | (0.067) | (0.053) | (0.093) | | (0.115) | (0.082) | (0.094) | (0.112) |
| | | Log Likelihood | -521.5 | -521.0 | -526.8 | | -528.2 | -529.0 | -520.5 | -527.3 |
| | | $\chi^2$ | 14.40 | 15.52 | 3.90 | | 0.97 | 0.06 | 16.36 | 2.82 |
| | | P-value [a] | 0.0007 | 0.0004 | 0.1426 | | 0.6145 | 0.9721 | 0.0003 | 0.2438 |
| | AM | Haplotype: 1 | -0.049 | 0.077 | 0.045 | | 0.012 | -0.012 | -0.066 | -0.057 |
| | | | (0.026) | (0.026) | (0.03) | | (0.031) | (0.028) | (0.029) | (0.034) |
| | | Haplotype: 2 | -0.069 | 0.133 | 0.192 | | -0.077 | -0.012 | -0.191 | -0.222 |
| | | | (0.067) | (0.055) | (0.09) | | (0.107) | (0.081) | (0.092) | (0.109) |
| | | Conditional Wald F-test [c] | 1.90 | 5.31 | 2.76 | | 0.42 | 0.10 | 3.77 | 3.00 |
| | | P-value [c] | 0.1494 | 0.0051 | 0.0638 | | 0.6598 | 0.9070 | 0.0234 | 0.0502 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (1 copy) | | | | | | | | |
| | | N (2 copies) | | | | | | | | |
| | LMM | Log Likelihood | -521.2 | -520.5 | | | | -526.0 | | |
| | | $\chi^2$ | 0.67 | 0.99 | | | | 5.51 | | |
| | | P-value [b] | 0.7151 | 0.6090 | | | | 0.0636 | | |
| | AM | Conditional Wald F-test [d] | 0.18 | 0.29 | | | | 1.66 | | |
| | | P-value [d] | 0.8376 | 0.7463 | | | | 0.1903 | | |

[a] compared to null model (log likelihood = -528.7)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.19. Model output for anti-*T. circ* IgG models of haplotype presence/absence in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1087 | 960 | 1260 | 1471 | 1261 | 1141 | 1178 | 1336 |
| | | N (present) | 488 | 615 | 315 | 104 | 314 | 434 | 397 | 239 |
| | LMM | Haplotype: 1 | -0.007 | -0.021 | 0.007 | 0.011 | 0.031 | -0.006 | -0.022 | 0.016 |
| | | | (0.010) | (0.009) | (0.012) | (0.019) | (0.012) | (0.010) | (0.011) | (0.013) |
| | | Log Likelihood | 462.2 | 464.4 | 462.1 | 462.1 | 465.5 | 462.1 | 464.0 | 462.7 |
| | | $\chi^2$ | 0.50 | 4.85 | 0.32 | 0.32 | 7.08 | 0.34 | 4.19 | 1.59 |
| | | P-value [a] | 0.4794 | 0.0276 | 0.5690 | 0.5710 | 0.0078 | 0.5612 | 0.0407 | 0.2069 |
| | AM | Haplotype: 1 | -0.002 | -0.018 | 0.007 | -0.007 | 0.029 | 0.001 | -0.029 | 0.008 |
| | | | (0.01) | (0.01) | (0.012) | (0.021) | (0.012) | (0.011) | (0.011) | (0.014) |
| | | Conditional Wald F-test [c] | 0.04 | 3.11 | 0.31 | 0.12 | 5.85 | 0.00 | 6.83 | 0.34 |
| | | P-value | 0.8441 | 0.0779 | 0.5776 | 0.7309 | 0.0157 | 0.9522 | 0.0091 | 0.5610 |
| With interaction | | N (absent) | F=552 M=535 | F=489 M=471 | F=639 M=621 | F=751 M=720 | F=646 M=615 | F=587 M=554 | F=609 M=569 | F=684 M=652 |
| | | N (present) | F=253 M=235 | F=316 M=299 | F=166 M=149 | F=54 M=50 | F=159 M=155 | F=218 M=216 | F=196 M=201 | F=121 M=118 |
| | LMM | Log Likelihood | 462.4 | 464.4 | 463.0 | 462.1 | 466.4 | 462.4 | 464.1 | 464.0 |
| | | $\chi^2$ | 0.42 | 0.15 | 1.73 | 0.01 | 1.76 | 0.53 | 0.10 | 2.49 |
| | | P-value [b] | 0.5158 | 0.7029 | 0.1884 | 0.9033 | 0.1840 | 0.4658 | 0.7519 | 0.1149 |
| | AM | Conditional Wald F-test [d] | 0.37 | 0.00 | 1.20 | 0.07 | 2.32 | 0.17 | 0.50 | 2.10 |
| | | P-value | 0.5424 | 0.9655 | 0.2732 | 0.7930 | 0.1279 | 0.6823 | 0.4814 | 0.1476 |

[a] compared to null model (log likelihood = 461.9)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.20. Model output for anti-*T. circ* IgG models of haplotype dosage in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1087 | 960 | 1260 | 1471 | 1261 | 1141 | 1178 | 1336 |
| | | N (1 copy) | 436 | 530 | 301 | 100 | 300 | 397 | 369 | 229 |
| | | N (2 copies) | 52 | 85 | 14 | 4 | 14 | 37 | 28 | 10 |
| | LMM | Haplotype: 1 | -0.009 | -0.022 | 0.005 | | 0.026 | -0.009 | -0.021 | 0.010 |
| | | | (0.010) | (0.010) | (0.012) | | (0.012) | (0.011) | (0.011) | (0.013) |
| | | Haplotype: 2 | 0.012 | -0.012 | 0.036 | | 0.155 | 0.026 | -0.032 | 0.163 |
| | | | (0.025) | (0.021) | (0.048) | | (0.048) | (0.030) | (0.036) | (0.056) |
| | | Log Likelihood | 462.5 | 464.5 | 462.3 | | 469.1 | 462.8 | 464.1 | 466.3 |
| | | $\chi^2$ | 1.15 | 5.07 | 0.71 | | 14.28 | 1.67 | 4.28 | 8.74 |
| | | P-value [a] | 0.5636 | 0.0792 | 0.7023 | | 0.0008 | 0.4329 | 0.1179 | 0.0126 |
| | AM | Haplotype: 1 | -0.004 | -0.018 | 0.006 | | 0.025 | -0.001 | -0.029 | 0.003 |
| | | | (0.011) | (0.01) | (0.012) | | (0.012) | (0.011) | (0.011) | (0.014) |
| | | Haplotype: 2 | 0.018 | -0.011 | 0.02 | | 0.157 | 0.031 | -0.042 | 0.151 |
| | | | (0.026) | (0.022) | (0.048) | | (0.047) | (0.03) | (0.036) | (0.055) |
| | | Conditional Wald F-test [c] | 0.37 | 1.61 | 0.20 | | 6.81 | 0.58 | 3.48 | 3.73 |
| | | P-value [c] | 0.6904 | 0.2000 | 0.8212 | | 0.0011 | 0.5590 | 0.0311 | 0.0242 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (1 copy) | | | | | | | | |
| | | N (2 copies) | | | | | | | | |
| | LMM | Log Likelihood | 462.8 | 464.7 | | | | 463.0 | 464.1 | |
| | | $\chi^2$ | 0.53 | 0.39 | | | | 0.50 | 0.11 | |
| | | P-value [b] | 0.7669 | 0.8217 | | | | 0.7802 | 0.9447 | |
| | AM | Conditional Wald F-test [d] | 0.26 | 0.08 | | | | 0.09 | 0.27 | |
| | | P-value [d] | 0.7728 | 0.9194 | | | | 0.9104 | 0.7629 | |

[a] compared to null model (log likelihood = -450.2)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.21. Model output for anti-*T. circ* IgG models of haplotype presence/absence in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 706 | 875 | 1076 | 915 | 844 | 869 | 954 |
| | | N (present) | 369 | 434 | 265 | 64 | 225 | 296 | 271 | 186 |
| | LMM | Haplotype: 1 | -0.033 | 0.014 | 0.021 | -0.042 | 0.033 | -0.043 | 0.014 | 0.024 |
| | | | (0.011) | (0.011) | (0.013) | (0.023) | (0.013) | (0.012) | (0.013) | (0.015) |
| | | Log Likelihood | 1134.3 | 1131.1 | 1131.6 | 1132.0 | 1133.3 | 1136.3 | 1130.9 | 1131.6 |
| | | $\chi^2$ | 8.06 | 1.69 | 2.71 | 3.44 | 6.12 | 12.09 | 1.29 | 2.72 |
| | | P-value [a] | 0.0045 | 0.1938 | 0.1000 | 0.0638 | 0.0134 | 0.0005 | 0.2566 | 0.0989 |
| | AM | Haplotype: 1 | -0.025 | 0.016 | 0.023 | -0.051 | 0.037 | -0.038 | 0.004 | 0.015 |
| | | | (0.012) | (0.012) | (0.014) | (0.026) | (0.015) | (0.013) | (0.014) | (0.016) |
| | | Conditional Wald F-test | 4.00 | 1.61 | 2.62 | 3.86 | 6.57 | 8.30 | 0.08 | 0.86 |
| | | P-value [c] | 0.0460 | 0.2052 | 0.1060 | 0.0501 | 0.0106 | 0.0041 | 0.7721 | 0.3549 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (present) | | | | | | | | |
| | LMM | Log Likelihood | 1137.2 | 1132.1 | 1131.7 | 1133.4 | 1133.4 | 1137.8 | 1131.9 | 1132.5 |
| | | $\chi^2$ | 5.96 | 2.13 | 0.20 | 2.92 | 0.15 | 3.09 | 1.99 | 1.88 |
| | | P-value [b] | 0.0146 | 0.1446 | 0.6512 | 0.0873 | 0.7013 | 0.0789 | 0.1584 | 0.1709 |
| | AM | Conditional Wald F-test | 5.81 | 1.54 | 0.08 | 2.80 | 0.04 | 2.97 | 1.58 | 2.32 |
| | | P-value [d] | 0.0161 | 0.2149 | 0.7770 | 0.0946 | 0.8441 | 0.0851 | 0.2091 | 0.1283 |

[a] compared to null model (log likelihood = 1130.2)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.22. Model output for anti-*T. circ* IgG models of haplotype dosage in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 771 | 706 | 875 | 1076 | 915 | 844 | 869 | 954 |
| | | N (1 copy) | 335 | 379 | 248 | 63 | 214 | 274 | 253 | 174 |
| | | N (2 copies) | 34 | 55 | 17 | 1 | 11 | 22 | 18 | 12 |
| | LMM | Haplotype: 1 | -0.032 | 0.019 | 0.024 | | 0.032 | -0.041 | 0.008 | 0.022 |
| | | | (0.012) | (0.012) | (0.013) | | (0.014) | (0.013) | (0.013) | (0.015) |
| | | Haplotype: 2 | -0.039 | -0.015 | -0.019 | | 0.053 | -0.060 | 0.116 | 0.054 |
| | | | (0.032) | (0.025) | (0.045) | | (0.055) | (0.039) | (0.045) | (0.054) |
| | | Log Likelihood | 1134.3 | 1132.0 | 1132.0 | | 1133.4 | 1136.4 | 1133.6 | 1131.8 |
| | | $\chi^2$ | 8.10 | 3.42 | 3.57 | | 6.26 | 12.30 | 6.77 | 3.06 |
| | | P-value [a] | 0.0174 | 0.1804 | 0.1679 | | 0.0437 | 0.0021 | 0.0339 | 0.2168 |
| | AM | Haplotype: 1 | -0.024 | 0.019 | 0.024 | | 0.036 | -0.038 | 0.001 | 0.013 |
| | | | (0.013) | (0.012) | (0.014) | | (0.015) | (0.013) | (0.014) | (0.016) |
| | | Haplotype: 2 | -0.028 | -0.022 | 0.004 | | 0.073 | -0.043 | 0.074 | 0.044 |
| | | | (0.033) | (0.027) | (0.045) | | (0.054) | (0.040) | (0.046) | (0.054) |
| | | Conditional Wald F-test [c] | 2.00 | 2.05 | 1.41 | | 3.52 | 4.16 | 1.31 | 0.58 |
| | | P-value [c] | 0.1358 | 0.1298 | 0.2451 | | 0.0300 | 0.0160 | 0.2712 | 0.5574 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (1 copy) | | | | | | | | |
| | | N (2 copies) | | | | | | | | |
| | LMM | Log Likelihood | 1137.3 | 1133.0 | | | | 1138.4 | | |
| | | $\chi^2$ | 6.06 | 2.02 | | | | 4.06 | | |
| | | P-value [b] | 0.0484 | 0.3651 | | | | 0.1314 | | |
| | AM | Conditional Wald F-test [d] | 3.07 | 0.67 | | | | 2.04 | | |
| | | P-value [d] | 0.0470 | 0.5123 | | | | 0.1305 | | |

[a] compared to null model (log likelihood = 1130.2)
[b] compared to model without interaction
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.23. Model output for FEC models of haplotype presence/absence in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1111 | 968 | 1290 | 1529 | 1285 | 1197 | 1220 | 1379 |
| | | N (present) | 509 | 652 | 330 | 91 | 335 | 423 | 400 | 241 |
| | NB | Haplotype: 1 | 0.113 | 0.115 | -0.136 | -0.058 | -0.011 | -0.121 | -0.066 | -0.092 |
| | | | (0.063) | (0.059) | (0.071) | (0.123) | (0.072) | (0.065) | (0.067) | (0.081) |
| | | Log Likelihood | -10656.7 | -10656.3 | -10656.5 | -10658.2 | -10658.3 | -10656.6 | -10657.8 | -10657.7 |
| | | P-value [a] | 0.0736 | 0.0455 | 0.0578 | 0.6547 | 1.0000 | 0.0652 | 0.3173 | 0.2733 |
| | AM | Haplotype: 1 | 0.084 | 0.069 | -0.093 | -0.076 | 0.019 | -0.104 | -0.001 | -0.031 |
| | | | (0.051) | (0.049) | (0.059) | (0.104) | (0.059) | (0.055) | (0.055) | (0.067) |
| | | Conditional Wald F-test [c] | 3.80 | 1.00 | 1.96 | 1.00 | 0.05 | 1.90 | 0.15 | 2.00 |
| | | P-value | 0.0225 | 0.3676 | 0.1415 | 0.3668 | 0.9513 | 0.1508 | 0.8607 | 0.1352 |
| With interaction | | N (absent) | | | | | | | | |
| | | N (present) | | | | | | | | |
| | NB | Log Likelihood | -10656.6 | -10656.3 | -10656.3 | -10658.1 | -10658.2 | -10656.2 | -10657.8 | -10657.5 |
| | | P-value [b] | 0.6547 | 1.0000 | 0.5271 | 0.6547 | 0.6547 | 0.3711 | 1.0000 | 0.5271 |
| | AM | Conditional Wald F-test [d] | 0.02 | 0.07 | 0.21 | 0.14 | 0.32 | 0.68 | 1.64 | 1.41 |
| | | P-value | 0.8758 | 0.7911 | 0.6503 | 0.7047 | 0.5693 | 0.4084 | 0.2011 | 0.2358 |

[a] compared to null model (log likelihood = -10658.3)
[b] compared to model without interaction
[c] Modelled using Log(FEC+50) with Gaussian error. Significance of haplotype term in model without haplotype*sex interaction
[d] Modelled using Log(FEC+50) with Gaussian error. Significance of haplotype*sex term

Table F.24. Model output for FEC models of haplotype dosage in lambs.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Without interaction | | N (absent) | 1111 | 968 | 1290 | 1529 | 1285 | 1197 | 1220 | 1379 |
| | | N (1 copy) | 451 | 559 | 316 | 87 | 322 | 381 | 374 | 232 |
| | | N (2 copies) | 58 | 93 | 14 | 4 | 13 | 42 | 26 | 9 |
| | NB | Haplotype: 1 | 0.044 | 0.119 | -0.128 | | -0.0075 | -0.123 | -0.066 | |
| | | | (0.065) | (0.061) | (0.072) | | (0.073) | (0.068) | (0.069) | |
| | | Haplotype: 2 | 0.591 | 0.090 | -0.317 | | -0.103 | -0.100 | -0.058 | |
| | | | (0.161) | (0.127) | (0.299) | | (0.346) | (0.178) | (0.240) | |
| | | Log Likelihood | -10650.5 | -10656.3 | -10656.3 | | -10658.3 | -10656.6 | -10657.8 | |
| | | P-value [a] | 0.0004 | 0.1353 | 0.1353 | | 1.0000 | 0.1827 | 0.6065 | |
| | AM | Haplotype: 1 | 0.057 | 0.07 | -0.082 | | 0.019 | -0.1 | -0.006 | |
| | | | (0.052) | (0.05) | (0.06) | | (0.059) | (0.056) | (0.056) | |
| | | Haplotype: 2 | 0.335 | 0.062 | -0.359 | | 0.009 | -0.159 | 0.097 | |
| | | | (0.125) | (0.104) | (0.229) | | (0.267) | (0.142) | (0.188) | |
| | | Conditional Wald F-test [c] | 2.70 | 2.00 | 2.47 | | 0.10 | 3.62 | 0.00 | |
| | | P-value [c] | 0.1008 | 0.1578 | 0.1162 | | 0.7533 | 0.0574 | 0.9828 | |
| With interaction | | N (absent) | | | | | | | | |
| | | N (1 copy) | | | | | | | | |
| | | N (2 copies) | | | | | | | | |
| | NB | Log Likelihood | -10648.8 | -10654.9 | -10656.1 | | -10657.3 | -10656.2 | -10657.8 | |
| | | P-value [b] | 0.1827 | 0.2466 | 0.8187 | | 0.3679 | 0.6703 | 1.0000 | |
| | AM | Conditional Wald F-test [d] | 0.01 | 0.72 | 0.10 | | 0.71 | 0.38 | 1.01 | |
| | | P-value [d] | 0.9913 | 0.4889 | 0.9053 | | 0.4926 | 0.6831 | 0.3661 | |

[a] compared to null model (log likelihood = -2103.5)

[b] compared to model without interaction

[c] Modelled using Log(FEC+50) with Gaussian error. Significance of haplotype term in model without haplotype*sex interaction

[d] Modelled using Log(FEC+50) with Gaussian error. Significance of haplotype*sex term

Table F.25. Model output for FEC models of haplotype presence/absence in yearlings.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Females | | N (absent) | 306 | 268 | 332 | 400 | 338 | 316 | 325 | 368 |
| | | N (present) | 127 | 165 | 101 | 33 | 95 | 117 | 108 | 65 |
| | NB | Haplotype: 1 | 0.222 | 0.052 | -0.103 | -0.390 | -0.056 | -0.045 | -0.121 | 0.000 |
| | | | (0.175) | (0.165) | (0.190) | (0.300) | (0.196) | (0.181) | (0.188) | (0.229) |
| | | Log Likelihood | -2624.1 | -2624.9 | -2624.8 | -2624.2 | -2624.9 | -2624.9 | -2624.7 | -2624.9 |
| | | P-value [a] | 0.1976 | 0.7518 | 0.5967 | 0.2207 | 0.7773 | 0.8065 | 0.5271 | 1.0000 |
| | AM | Haplotype: 1 | 0.179 | 0.078 | -0.068 | -0.28 | -0.052 | -0.031 | -0.133 | -0.005 |
| | | | (0.097) | (0.094) | (0.11) | (0.174) | (0.11) | (0.101) | (0.106) | (0.127) |
| | | Conditional Wald F-test | 3.40 | 0.69 | 0.38 | 2.60 | 0.23 | 0.10 | 1.58 | 0.00 |
| | | P-value | 0.0659 | 0.4072 | 0.5406 | 0.1080 | 0.6353 | 0.7563 | 0.2104 | 0.9701 |
| Males | | N (absent) | 216 | 183 | 243 | 292 | 254 | 230 | 232 | 244 |
| | | N (present) | 92 | 125 | 65 | 16 | 54 | 78 | 76 | 64 |
| | NB | Haplotype: 1 | 0.300 | -0.101 | 0.198 | 0.088 | -0.245 | -0.114 | -0.304 | 0.174 |
| | | | (0.149) | (0.141) | (0.166) | (0.301) | (0.177) | (0.158) | (0.161) | (0.173) |
| | | Log Likelihood | -2071.1 | -2072.9 | -2072.4 | -2073.1 | -2072.2 | -2072.9 | -2071.4 | -2072.6 |
| | | P-value [b] | 0.0419 | 0.4708 | 0.2238 | 0.7518 | 0.1773 | 0.4708 | 0.0652 | 0.3078 |
| | AM | Haplotype: 1 | 0.297 | -0.013 | 0.196 | 0.134 | -0.274 | -0.113 | -0.25 | 0.172 |
| | | | (0.115) | (0.109) | (0.131) | (0.237) | (0.135) | (0.124) | (0.126) | (0.133) |
| | | Conditional Wald F-test | 6.61 | 0.01 | 2.23 | 0.32 | 4.13 | 0.84 | 3.90 | 1.65 |
| | | P-value | 0.011 | 0.906 | 0.137 | 0.573 | 0.043 | 0.361 | 0.049 | 0.200 |

[a] compared to null model (log likelihood = -2624.9)
[b] compared to null model (log likelihood = -7073.12)
[c] significance of haplotype term in model without haplotype*sex interaction
[d] significance of haplotype*sex term

Table F.26. Model output for FEC models of haplotype presence/absence in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Females | | N (absent) | 416 | 390 | 464 | 581 | 497 | 464 | 471 | 533 |
| | | N (present) | 204 | 230 | 156 | 39 | 123 | 156 | 149 | 87 |
| | NB | Haplotype: 1 | -0.066 | 0.048 | 0.050 | 0.202 | -0.136 | 0.118 | -0.119 | -0.178 |
| | | | (0.14) | (0.133) | (0.154) | (0.242) | (0.162) | (0.146) | (0.146) | (0.194) |
| | | Log Likelihood | 0.22 | 0.12 | 0.10 | 0.74 | 0.68 | 0.66 | 0.80 | 0.22 |
| | | P-value [a] | 0.6390 | 0.7290 | 0.7518 | 0.3897 | 0.4096 | 0.4166 | 0.3711 | 0.6390 |
| | AM | Haplotype: 1 | -0.033 | 0.029 | 0.044 | 0.186 | -0.035 | 0.038 | -0.088 | -0.055 |
| | | | (0.051) | (0.051) | (0.057) | (0.096) | (0.059) | (0.054) | (0.056) | (0.069) |
| | | Conditional Wald F-test | 0.44 | 0.32 | 0.58 | 3.73 | 0.36 | 0.48 | 2.46 | 0.65 |
| | | P-value | 0.5092 | 0.5702 | 0.4478 | 0.0546 | 0.5486 | 0.4869 | 0.1180 | 0.4223 |
| Males | | N (absent) | 159 | 146 | 186 | 218 | 193 | 163 | 179 | 199 |
| | | N (present) | 76 | 89 | 49 | 17 | 42 | 72 | 56 | 36 |
| | NB | Haplotype: 1 | 0.023 | 0.213 | -0.363 | -0.070 | -0.145 | 0.218 | -0.052 | -0.048 |
| | | | (0.133) | (0.126) | (0.156) | (0.222) | (0.146) | (0.127) | (0.144) | (0.173) |
| | | Log Likelihood | 0.04 | 2.90 | 5.04 | 0.10 | 0.96 | 2.98 | 0.14 | 0.08 |
| | | P-value [b] | 0.8415 | 0.0886 | 0.0248 | 0.7518 | 0.3272 | 0.0843 | 0.7083 | 0.7773 |
| | AM | Haplotype: 1 | 0.104 | 0.131 | -0.355 | -0.015 | -0.218 | 0.242 | 0.01 | -0.113 |
| | | | (0.102) | (0.099) | (0.119) | (0.179) | (0.117) | (0.101) | (0.113) | (0.134) |
| | | Conditional Wald F-test | 1.04 | 1.77 | 8.86 | 0.01 | 3.44 | 5.78 | 0.01 | 0.71 |
| | | P-value | 0.3090 | 0.1852 | 0.0032 | 0.9312 | 0.0653 | 0.0172 | 0.9301 | 0.4003 |

[a] compared to null model (log likelihood = -7951.4)
[b] compared to null model (log likelihood = -3021.5)

Table F.27. Model output for FEC models of haplotype dosage in adults.

| | | | Haplotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Females | | N (absent) | 416 | 390 | 464 | 581 | 497 | 464 | 471 | 533 |
| | | N (1 copy) | 187 | 193 | 146 | 38 | 119 | 145 | 138 | 82 |
| | | N (2 copies) | 17 | 37 | 10 | 1 | 4 | 11 | 11 | 5 |
| | NB | Haplotype: 1 | -0.083 | 0.022 | -0.003 | | -0.114 | 0.097 | -0.087 | |
| | | | (0.144) | (0.143) | (0.157) | | (0.165) | (0.149) | (0.149) | |
| | | Haplotype: 2 | 0.132 | 0.151 | 0.735 | | -0.834 | 0.456 | -0.817 | |
| | | | (0.459) | (0.253) | (0.608) | | (0.693) | (0.566) | (0.501) | |
| | | Log Likelihood | 0.36 | 1.86 | 0.74 | | 1.10 | 2.32 | 1.12 | |
| | | P-value [a] | 0.8353 | 0.3946 | 0.6907 | | 0.5769 | 0.3135 | 0.5712 | |
| | AM | Haplotype: 1 | -0.05 | 0.013 | 0.028 | | -0.027 | 0.018 | -0.059 | |
| | | | (0.052) | (0.053) | (0.059) | | (0.06) | (0.055) | (0.057) | |
| | | Haplotype: 2 | -0.051 | 0.051 | 0.513 | | -0.158 | 0.192 | -0.43 | |
| | | | (0.156) | (0.099) | (0.203) | | (0.271) | (0.195) | (0.179) | |
| | | Conditional Wald F-test | 0.47 | 0.14 | 3.22 | | 0.26 | 0.51 | 3.11 | |
| | | P-value | 0.624 | 0.870 | 0.041 | | 0.774 | 0.599 | 0.046 | |
| Males – N | | N (absent) | 159 | 146 | 186 | 218 | 193 | 163 | 179 | 199 |
| | | N (1 copy) | 71 | 82 | 44 | 17 | 39 | 64 | 54 | 33 |
| | | N (2 copies) | 5 | 7 | 5 | 0 | 3 | 8 | 2 | 3 |
| | NB | Haplotype: 1 | -0.010 | 0.234 | | | | 0.267 | | |
| | | | (0.137) | (0.13) | | | | (0.134) | | |
| | | Haplotype: 2 | 0.309 | -0.066 | | | | -0.139 | | |
| | | | (0.374) | (0.359) | | | | (0.278) | | |
| | | Log Likelihood | 0.78 | 3.54 | | | | 4.80 | | |
| | | P-value [b] | 0.6771 | 0.1703 | | | | 0.0907 | | |
| | AM | Haplotype: 1 | 0.073 | 0.138 | | | | 0.277 | | |
| | | | (0.106) | (0.101) | | | | (0.105) | | |
| | | Haplotype: 2 | 0.451 | 0.05 | | | | -0.001 | | |
| | | | (0.305) | (0.293) | | | | (0.236) | | |
| | | Conditional Wald F-test | 1.243 | 0.923 | | | | 0.711 | | |
| | | P-value | 0.291 | 0.399 | | | | 0.492 | | |

[a] compared to null model (log likelihood = -7951.4)
[b] compared to null model (log likelihood = -3021.5)

Table F.28. Associations between haplotype number and each of the five traits in each age-sex category for the GLMMs and AMs.

| | | GLMM | | AM | |
|---|---|---|---|---|---|
| Trait | Age-sex category | $\chi^2$ | P value | Conditional Wald F-test | P-value |
| August weight | Lambs | 1.230 | 0.2674 | 0.196 | 0.658 |
| | Yearlings | 0.024 | 0.8761 | 0.149 | 0.700 |
| | Adults | 0.612 | 0.4341 | 1.171 | 0.279 |
| Anti-*T. circ* IgA | Lambs | 1.925 | 0.6019 | 0.272 | 0.272 |
| | Adults | 1.469 | 0.2255 | 0.085 | 0.771 |
| Anti-*T. circ* IgE | Lambs | 0.100 | 0.7520 | 0.066 | 0.798 |
| | Adults | 0.007 | 0.9330 | 0.059 | 0.807 |
| Anti-*T. circ* IgG | Lambs | 3.606 | 0.0576 | 3.344 | 0.068 |
| | Adults | 0.808 | 0.3687 | 0.389 | 0.533 |
| FEC | Lambs | | 0.040 | 0.40 | 0.529 |
| | Yearling females | | 0.390 | 0.82 | 0.366 |
| | Yearling males | | 0.888 | 0.32 | 0.572 |
| | Adult females | | 0.655 | 0.06 | 0.804 |
| | Adult males | | 0.729 | 0.01 | 0.936 |