A rare MHC haplotype confers selective advantage in a free-living ruminant

Wei Huang, Kara L Dicks, Jarrod D Hadfield, Susan E Johnston, Keith T Ballingall, Josephine M Pemberton

Abstract

The major histocompatibility complex (MHC) is the most variable gene family known in vertebrates. Parasite-mediated selection (PMS) is believed to be the main force maintaining diversity at MHC genes, but it has proven hard to demonstrate the exact PMS regime that is acting in natural populations. Demonstrating contemporary selection on MHC alleles is not trivial, and previous work has been constrained by limited genetic tools, low sample sizes, short time scales and anticonservative statistical approaches. Here, we use a conservative statistical approach to examine associations between MHC genes and fitness components, using haplotypes of expressed MHC class II genes in a large sample of Soay sheep *(Ovis aries*) monitored over their lifetimes between 1989 and 2012. Of the eight MHC haplotypes (A-H) identified, we found that haplotype C was associated with decreased male breeding success, haplotype D was associated with increased female life span, and haplotype F was associated with decreased female life span. Consistent with the increased lifespan in females, haplotype D has increased in frequency throughout the study period. Our results suggest the existence of contemporary balancing selection on MHC class II genes in Soay sheep as might be expected under the hypotheses of rare allele advantage or fluctuating selection and do not support an effect of heterozygous advantage.

Introduction

The vertebrate immune system has evolved to protect the host from infection through a variety of innate and adaptive immune mechanisms. Many of the proteins that control the specificity of immune function are the products of genes that are amongst the most diverse within vertebrate populations. Such diversity is believed to have evolved in response to pathogen diversity. Investigating the mechanisms shaping the diversity of immune genes has been a major focus in evolutionary biology (1). Genes located within the major histocompatibility complex (MHC) encode proteins that play important roles in the vertebrate immune system and have been intensively studied across many species, including model organisms (2-4). MHC molecules, the products of MHC class I and class II genes, present peptide fragments to antigen-specific receptors on T cells to induce adaptive immune responses. MHC class I molecules present peptides derived from intracellular pathogens such as viruses to CD8 +ve T cells; while MHC class II molecules present peptides derived from extracellular pathogens such as bacteria and parasites to CD4 +ve T cells (5).

MHC class I and II genes are the most polymorphic loci identified in vertebrate populations with hundreds of alleles described in many species (6). Pathogen-mediated balancing selection is thought to be the principal driver maintaining such high levels of allelic diversity (7-10). However, the specific selection mechanisms which act to maintain high levels of MHC allelic diversity are not clear (3, 11, 12). Three models have been proposed to explain how PMS acts to maintain such diversity:

1) *Heterozygote advantage*: individuals which are more heterozygous across the MHC can respond to a greater variety of pathogens. i.e. there is a consistent advantage to heterozygosity regardless of the alleles involved. As a result, variation at MHC loci is selected and maintained (13).

2) *Rare-allele advantage*: there is strong selection on pathogens to evade immune protection provided by the most common host MHC alleles. Therefore, rare alleles have a selective advantage. However, the advantage disappears as the frequency of a protective MHC allele increases and is absent at equilibrium. The arms race between hosts and pathogens sets up a cyclical process causing MHC alleles and pathogens to fluctuate, meaning that MHC diversity is maintained by negative frequency-dependent selection (14). It should be noted that

since rare alleles appear most frequently as heterozygotes, this can manifest as heterozygote advantage for specific alleles under some circumstances.

3) *Fluctuating selection*: a changing environment results in variation in the abundance of different pathogens in space and time, which generates directional selection in different host subpopulations and/or at different time points. Thus, different alleles will be favoured in different subpopulations or at different times, meaning MHC diversity is maintained as a consequence (15, 16). Unlike rare-allele advantage the change in selection coefficients in time and/or space is not driven by coevolution in the pathogens.

To test each selection model we can examine associations between MHC diversity and parasite presence or parasite load; different patterns of associations can be explained as a result of different selection mechanisms. An association between MHC heterozygosity and reduced parasite load could be viewed as evidence of heterozygous advantage, while if MHC alleles or haplotypes are correlated with protection against a particular parasite species, rare-allele advantage or fluctuating selection could be the mechanism (12). A number of studies on MHC genes in natural populations have examined associations between MHC variation and parasites, often with conflicting results. Some studies found that MHC heterozygosity or certain MHC alleles confer resistance to parasite infection (17-19), while another study found no association between MHC alleles and parasite load (20). However, prior to the development of high throughput sequencing technologies, such studies were probably limited by incomplete data on the diversity and abundance of the parasite populations present. As a consequence, parasites that strongly select for specific MHC alleles or haplotypes may have been missed. Therefore, quantifying relationships between MHC alleles and pathogens may not have been able to reliably elucidate which selective mechanisms were operating. An alternative means to test for selection on MHC genes would be to examine MHC-dependent fitness patterns (4, 21).

As with studies examining associations between MHC and parasites, MHC-fitness associations can also indicate different selection mechanisms. Under heterozygote advantage, heterozygotes are always fitter regardless of the alleles they carry, so we predict a positive association between MHC heterozygosity and fitness. Under rare-allele advantage, if MHC alleles have reached equilibrium frequency, no association between MHC alleles and fitness should be detectable. However, when MHC alleles are not in equilibrium, we may detect selection favouring specific rare alleles and probably also disfavouring specific common alleles. Under fluctuating selection, we would also observe specific alleles being favoured or disfavoured by selection. But unlike rare-allele advantage, this should occur regardless of their frequency. However, it is

hard to distinguish rare-allele advantage and fluctuation selection by examining MHC-fitness associations (12). Because measuring fitness is generally challenging except in some long-term individual-based studies of natural populations, only a few studies have tested associations between fitness and MHC variation in the wild. Moreover, most such studies have only tested associations between MHC and a single fitness component, again with mixed results. First, some MHC studies in wild populations have found a positive association between MHC heterozygosity and fitness. For instance, the number of MHC alleles was found to be positively associated with increased apparent survival (return rate from initial year of capture) in common yellowthroats *(Geothlypis trichas)* (22). Similarly, Banks et al. (23) found there was an association between MHC heterozygosity and survival in a natural population of mountain brushtail possums *(Trichosurus cunninghami)*. Second, some studies have reported associations between fitness and specific MHC alleles or supertypes (clusters of MHC alleles with similar physicochemical properties at their antigen-binding sites). For example, Lukasch et al. (24) found that a specific MHC allele was associated with higher survival of house sparrow fledglings *(Passer domesticus)*. Another study of great tits *(Parus major)* using mark-recapture data identified three MHC supertypes associated with fitness components, one with increased survival, one with increased lifetime reproductive success and another one with decreased lifetime reproductive success (25). Kloch et al. (26) also found that two MHC alleles from different clusters were associated with winter survival in root voles *(Microtus oeconomus)*. Finally, there are some studies where no associations are detected. In collared flycatcher (*Ficedula albicollis*), no association between MHC genes and lifetime reproductive success were found (20). Similar in house sparrow (*Passer domesticus*), no significant association between MHC genes and survival were identified (27).

Many studies of selection on the MHC are also limited by the depth and quality of the genetic data available. Within a typical mammalian MHC region, multiple closely linked and duplicated genes are present which are inherited together as a haplotype. With the development of next generation sequencing (NGS) technologies, we are now able to cost-effectively genotype large numbers of samples (28, 29). However, in species where the MHC region is poorly characterised such as in fish or passerine birds, it is difficult to assign a sequence to an individual locus or identify if a locus is functional. Without locus-specific genotyping, multiple loci may co-amplify, which will impact subsequent testing for associations between MHC diversity and fitness. Even if locus-specific genotyping is possible, most studies have focused on a single locus and ignored linkage with other loci (23, 26, 30). Locus-specific genotyping and characterisation of haplotypes is desirable when testing for associations between fitness and MHC diversity in wild populations.

Lastly, statistical analyses of selection on MHC genes require greater sophistication than is often deployed. Measures of parasite load and fitness are quantitative traits likely to be determined by both environmental factors and multiple genetic factors, including MHC genes. As such, it is possible that many studies of contemporary selection confound MHC similarity with genetic relatedness, so an MHC effect could be due to loci elsewhere in the genome rather than the MHC itself. Selection on MHC genes in natural populations has mostly been assessed in long-term individual-based studies which include a lot of genetically related individuals (25, 31). To our knowledge, no MHC study in a natural population has been conducted in a quantitative genetic framework and thus they are anti-conservative. Similarly, given the ubiquity of inbreeding depression, any analysis of selection on MHC testing for heterozygote advantage should include an estimate of individual inbreeding to ensure MHC effects are not confounded with genome-wide inbreeding effects.

Here, we present the first investigation of contemporary selection acting on the MHC in a free-living population that aims to overcome the issues raised above. The Soay sheep *(Ovis aries*) population on the island of Hirta, St Kilda is one of the most intensively studied wild animal populations in the world (32). Since 1985, individual Soay sheep living in the Village Bay study area have been followed from birth, through all breeding attempts, until death. In addition, using a combination of observation and SNP genotypes, a multigenerational pedigree for the population has been assembled which includes annual and lifetime measures of fitness for both sexes (33). Previous work characterised diversity at the MHC class II locus *DRB1* using an internal microsatellite and identified associations between MHC genotype, fitness and parasite infection (34, 35). However, the only fitness measure studied was juvenile survival. Another study using the same genetic dataset of MHC-linked microsatellite alleles found higher temporal and spatial variation in these loci than in putatively neutral microsatellite loci in other parts of the genome, suggesting fluctuating selection on MHC genes (36). Recently, using sequencing-based genotyping of functional MHC class II DR and DQ loci we have shown that eight MHC haplotypes are segregating in the study population (37). Using a panel of 13 SNPs, MHC class II haplotypes were characterised for 5349 individuals sampled between 1989 and 2012 (38). While diplotypes were consistently in Hardy-Weinberg equilibrium, haplotype frequencies are more even than expected by chance across all life history stages and standing population samples (Ewens-Watterson test), which is consistent with balancing selection (38). Here, we use this haplotype data to investigate associations between (1) MHC heterozygosity and fitness and (2) individual MHC haplotypes and fitness. We fit MHC haplotype within quantitative genetic models in order to distinguish the effect of MHC haplotypes from genome-wide effects. In addition (3), we perform gene-drop simulations to examine whether haplotype frequencies have changed more than expected from a neutral process.

Methods

Study system

Soay sheep have lived virtually unmanaged on the island of Soay, in the St. Kilda archipelago, for thousands of years. In 1932, 107 Soay sheep were introduced to the larger neighbouring island of Hirta and have been living there unmanaged since. From 1985, a long-term study has been conducted on the sheep resident in the Village Bay area of Hirta to investigate ecological and evolutionary questions (32). The vast majority of lambs are born in April or May of each year. At their first live capture, lambs are ear-tagged to enable life-long identification and discs of ear tissue removed in the tagging process are used for DNA extraction. Parentage is inferred for each individual using a subset of 315 unlinked SNPs derived from the Illumina Ovine 50K SNP array, on which most individuals alive since 1990 have been genotyped (39). In addition, the R package Sequoia (40) was used to cluster half-siblings which share a parent that has not been genotyped. In cases where no SNP genotypes were available, a small number of parentage inferences were made using field observations (for mothers) or a previous microsatellite genotype approach (41). The pedigree used here includes 7447 individuals, with 7014 paternal links and 6229 maternal links. This study investigated associations between MHC haplotypes and five fitness components defined as follows:

- Juvenile survival (JS), a binary variable indicating whether a lamb survived from birth until May 1st of the following year.
- Annual Survival (AS), a binary variable indicating whether an individual survived or not in a given calendar year;
- Life-span (LS), the age of an individual when it died;
- Annual Breeding Success (ABS), the number of offspring an individual produced in a given calendar year;
- Lifetime Breeding Success (LBS), the number of offspring an individual produced over its lifetime;

Analyses of LBS, ABS, LS and AS only included sheep that survived the juvenile period.

MHC Class IIa Haplotyping

The genetic data used in this study were obtained from a previous study (37, 38). In that study the seven expressed loci (*DRB1*, *DQA1*, *DQA2*, *DQA2-like*, *DQB1*, *DQB2* and *DQB2-like*) within the MHC class IIa region were characterised in 118 selected Soay sheep using sequencing-based genotyping. A total of eight MHC haplotypes were identified and named A to H. These haplotypes were confirmed in an additional 94

144 individuals that were unrelated based on the pedigree. A panel of 13 SNPs located in the region of MHC class IIa haplotypes was then selected
145 to impute the eight MHC class IIa haplotypes in 6752 Soay sheep genotyped using Kompetitive Allele-specific PCR (KASP). This panel included
146 11 SNPs known to be variable in Soays based on data from the Illumina Ovine infinium HD chip (42) and two other SNPs identified within *DQA1*
147 gene (42). After quality control, the diplotypes of 5349 individuals that lived in the study area between 1985 and 2012 were identified. As we
148 investigate cohort effects in the statistical analyses, we excluded animals born before 1989 since too few were genotyped to be representative
149 of their birth cohort; we also removed the cohort born in 2010 which had a high genotyping failure rate in the KASP genotyping assay. We also
150 excluded any individuals previously subjected to experimental treatments which may have affected their survival or breeding performance (e.g.
151 anthelmintic boluses). After these exclusions, a total of 3440 individuals (1750 males and 1650 females) remained for statistical modelling of JS
152 (Table 1). The frequency of rarest haplotype (D) was 3.02%. With the exception of animals homozygous for haplotype D (N=5), moderate
153 numbers of individuals (N>20) of each diplotype (i.e. haplotype complement) are represented (Supplementary 1).
154

155 # Statistical analysis

156 We used generalized linear mixed models (GLMMs) to study the associations between MHC haplotypes and fitness components. The primary
157 models were animal models (AMs) including a covariance structure proportional to the pedigree relatedness matrix, but we also repeated all
158 tests without pedigree information. In null models, we fitted a list of fixed and random effects for each fitness component (Table 1) including
159 genomic inbreeding (F_{grm}) as a fixed effect, as a previous study demonstrated inbreeding depression in this population (43). None of the five
160 fitness components in our study had a normal distribution, so we used Markov chain Monte Carlo techniques implemented in MCMCglmm v
161 2.28 (44) in R v3.5.2 (45) to run models. As the fitness components have different ontogenies and distributions between males and females
162 (Supplementary 2), separate models were built for males and females. For both males and females, we used probit regression for JS and AS
163 and we used Poisson regression for LS and LBS. As females can only have up to two offspring each year while males had between 0 and 24, we
164 used multiple category probit regression for female ABS and Poisson regression for male ABS. To examine whether there were associations
165 between MHC genes and fitness components, we fitted all haplotypes simultaneously into the null model. Each haplotype was fitted either as
166 a count (0, 1 or 2 copies) in the dosage models or as 1/0 in the presence/absence models. These two models capture the extreme scenarios of
167 pure additive gene action and complete dominance respectively. We included MHC heterozygosity as a fixed effect in both dosage and
168 presence/absence models, which assumes a common dominance deviation across MHC diplotypes. We performed a Wald test using the
169 posterior mean and posterior covariance matrix to test whether haplotypes explained variation in fitness. In these models, Haplotype H was

treated as a reference and dropped in our results so the results for individual haplotypes from our models are relative to haplotype H. Therefore, when a Wald test for haplotype differences was significant, we conducted an additional analysis (Supplementary 3) comparing the estimate for each haplotype with the mean of the estimates for all other haplotypes to identify individual haplotype effects.

Table 1. Fixed and random effects fitted in null model and sample sizes of each fitness component. Sample sizes are shown by number of individuals (records).

Fitness component	Fixed effects			Random effects			Sample sizes	
	Litter Size	Age (years, quadratic)	Inbreeding (F_{grm})	Birth Year	Year of Fitness Measure	ID	Female	Male
JS	×		×	×			1650	1750
AS		×	×	×	×	×	551 (3093)	529 (1610)
LS			×	×			574	545
ABS		×	×	×	×	×	574 (3667)	545 (2155)
LBS			×	×			574	545

Gene-drop analysis

We performed gene-drop simulations with a custom script in R v3.5.2 (45) to model the expected frequency change of MHC haplotypes under genetic drift given the pedigree structure. Hereafter, the first four birth cohorts (1989 to 1992) are defined as the "founder" cohorts, whereas all subsequent cohorts are defined as "simulated" cohorts (1993 to 2012). A total of 10,000 simulations were conducted as follows. Individuals in the founder cohorts were assigned their observed diplotypes; founder cohort individuals with unknown diplotypes were assigned one by either: (a) sampling a haplotype from each parent assuming Mendelian segregation; or (b) sampling a haplotype with the probability of the

184 observed haplotype frequencies in the same birth cohort when one or more parents were unknown. In the simulated cohorts, each individual
185 diplotype was sampled using approaches (a) and (b) above. By this method, simulated diplotypes were generated for each individual in the
186 pedigree born after 1992. Using these data, we compared the observed and simulated changes in MHC haplotype frequency for individuals
187 with known diplotypes, using a linear regression with cohort year (1993 to 2012) as the predictor variable. In cases where the observed slope
188 fell within the top or bottom 2.5% of simulated slopes, the allele frequency was deemed to have been subject to directional selection.
189

Results

Associations between MHC and fitness components

The results of models with and without pedigree information were largely similar (Supplementary 4 to 6). Here we report the results of animal models, i.e. including the pedigree as a random effect.

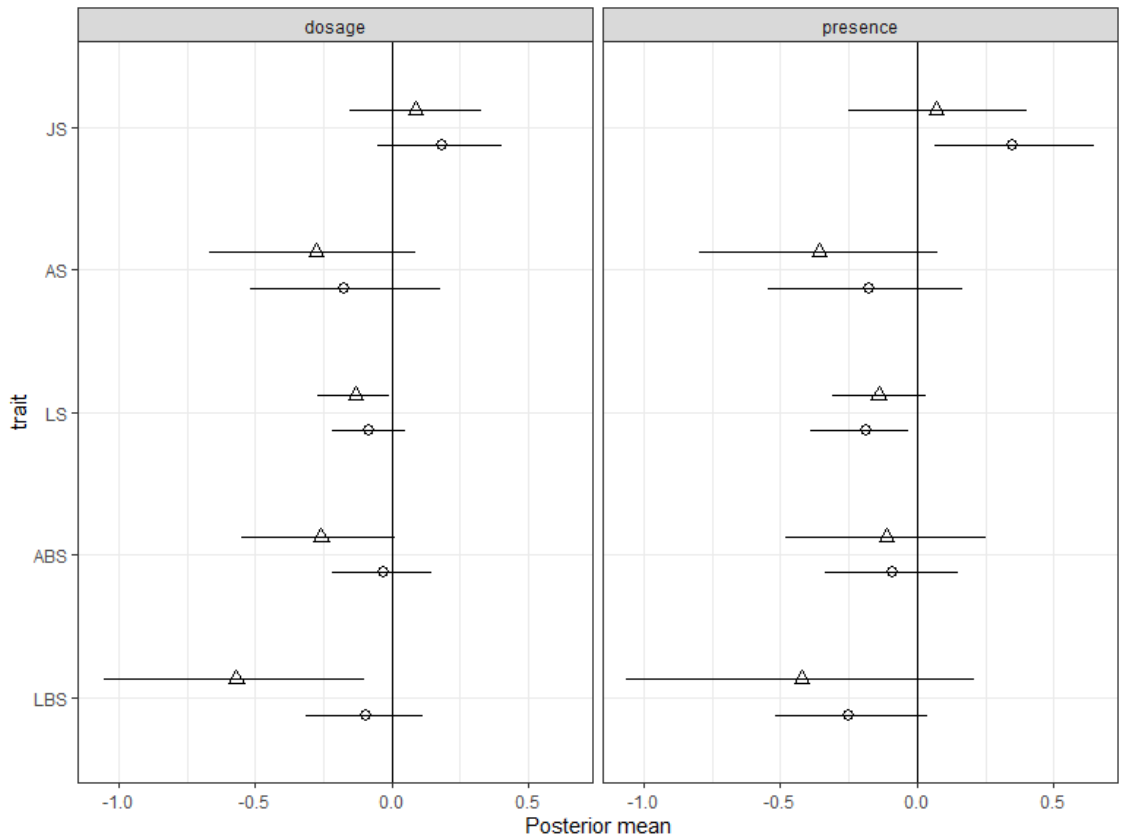

Figure 1. Associations between MHC heterozygosity and fitness components in Soay sheep in the dosage and presence/absence animal models in females (circles) and males (triangles). Bars represent the 95% credibility intervals and those not overlapped with zero represent significant results.

Table 2. Results (p value) of Wald tests for animal models (d.f.=7). Bold numbers show significant results (p <0.05).

Fitness component	Female		Male	
	Dosage model	Presence/absence model	Dosage model	Presence/absence model
JS	0.47	0.35	0.46	0.49
AS	0.068	0.066	0.43	0.56
LS	**0.039**	**0.046**	0.54	0.62
ABS	0.99	0.98	**0.024**	0.069
LBS	0.16	0.16	**0.044**	0.071

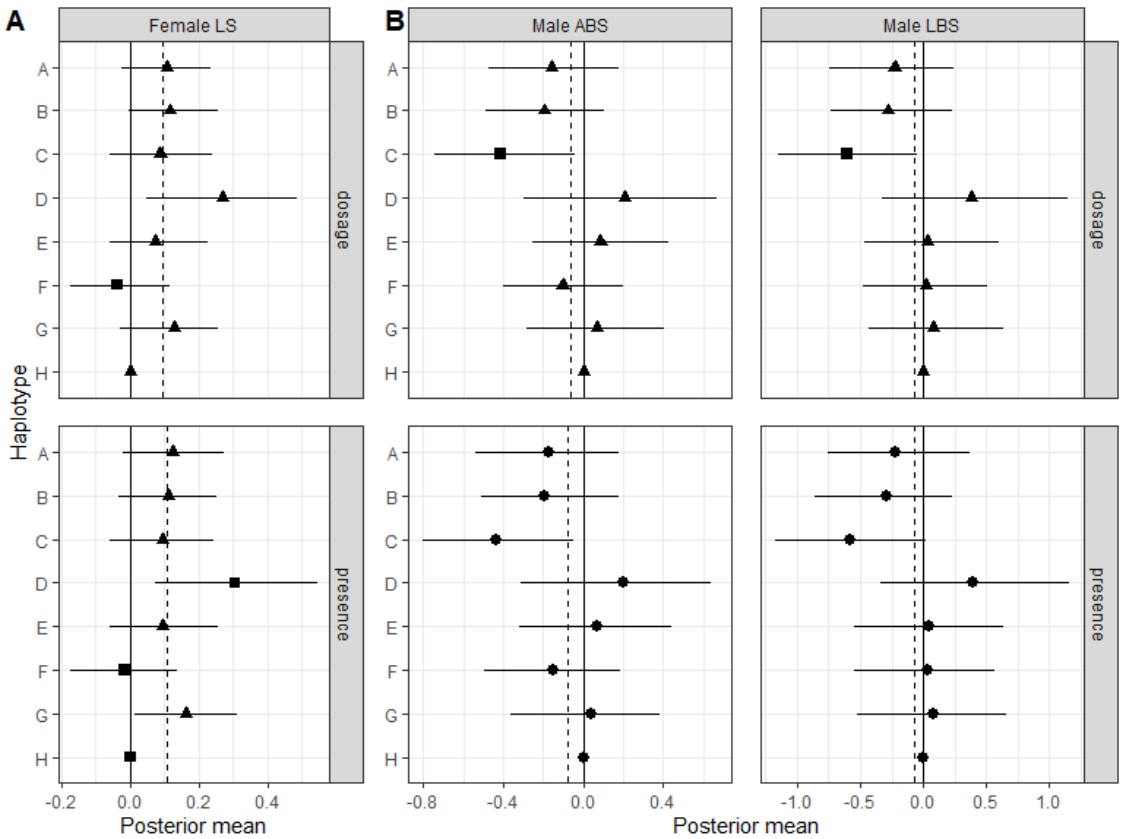

Figure 2. Associations between MHC haplotypes and fitness components in Soay sheep derived from animal models in females (A) and males (B). Only fitness traits in which Wald tests were significant in at least either dosage or presence/absence model are shown here. Bars represent the 95% credibility intervals. The solid lines represent the model intercept at haplotype H and posterior means and credible intervals for each haplotype are plotted relative to H. Dashed lines indicate the average posterior mean of all eight haplotypes. Circles showed that Wald test was not significant for this model. For models that Wald tests were significant, haplotype effects that were significantly different from the mean are presented in squares with non-significant ones presented in triangles.

Juvenile survival

We found a positive association between MHC heterozygosity and female JS in the presence/absence model (Figure 1). Considering haplotype effects, the Wald test was not significant in either the dosage or presence/absence models for either sex (Table 2, Supplementary Figure 4.2).

Annual survival

There was no association between MHC heterozygosity and annual survival (AS) in either sex (Figure 1). The Wald test for haplotype effects in AS was not significant in either the dosage or presence/absence models for either sex (Table 2, Supplementary Figure 4.2).

Lifespan

There was a negative association between MHC heterozygosity and male life span (LS) in the dosage model and a negative association between MHC heterozygosity and female life span (LS) in the presence/absence model (Figure 1). In the Wald test for haplotype effects a significant association was found for female LS in both dosage model and presence/absence models, but neither test was significant for males (Table 2, Figure 2A). Haplotype D was associated with increased female life span in the presence/absence model. Haplotype H was associated with decreased female life span in the dosage model while haplotype F was associated with decreased female life span in both the dosage and presence/absence models (Figure 2A).

Annual breeding success

There was no association between MHC heterozygosity and annual breeding success (ABS) in either sex (Figure 1). In the Wald test for haplotype effects, a significant association was found for male ABS in the dosage but not the presence /absence model; neither test was significant in female ABS (Table 2). Haplotype C was associated with decreased male ABS in the dosage model (Figure 2B).

Lifetime breeding success

There was a negative association between MHC heterozygosity and male lifetime breeding success (LBS) in the dosage model (Figure 1). The Wald test for haplotype effects was significant in male LBS but only in the dosage model; neither test was significant in female LBS (Table 2). Haplotype C was associated with decreased male LBS in the dosage model (Figure 2B).

The results obtained from models not including the pedigree were largely consistent with those above for associations between MHC heterozygosity and fitness components (Supplementary 4). For associations between individual MHC haplotypes and fitness components, estimates were largely similar but there were minor variations in terms of significance of Wald tests (Table 1, Supplementary 4). The Wald test was no longer significant for male LBS in the dosage model but the Wald test was significant for male ABS in the presence/absence model. Haplotype C was still associated with decreased male ABS in the presence/absence model (Supplementary 4).

Gene-drop analysis

All haplotype frequency changes were within expectation due to drift alone except for haplotype D (Figure 3, Supplementary 7). Of the 10000 simulations for Haplotype D, only 272 simulated regression slopes were greater than the slope generated from real data (p=0.028). However, this haplotype frequency change of haplotype was not significant for a two-tailed test.

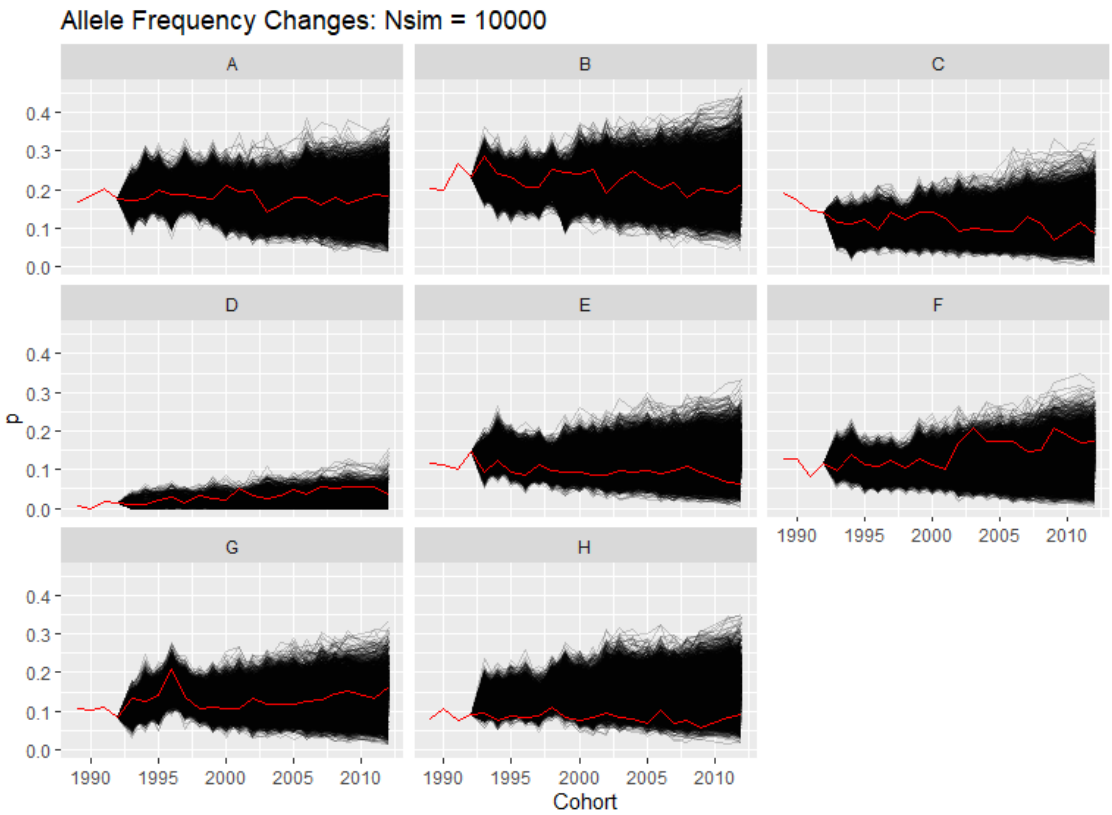

Figure 3. Results of the gene drop analysis for each haplotype. The black lines represent the result of simulations (N=10000) and the red lines represent the observed frequency of each haplotype in the Soay sheep population over the same time interval.

Discussion

In this study we investigated associations between MHC class II genes (heterozygosity and specific haplotypes) and five fitness components in a sample of individually-monitored free-living ruminants using generalized linear mixed models. We fitted individual MHC haplotype effects as both dosage and presence/absence and included MHC heterozygosity in all models. In addition, we included genetic relatedness to eliminate potential false positive results due to other genomic regions in common in related individuals. We found some associations between MHC heterozygosity and fitness components but the associations differed somewhat between dosage and presence/absence models. MHC heterozygosity was associated with increased female juvenile survival and decreased female LS in the presence/absence but not the dosage models; and MHC heterozygosity was associated with decreased male LS and LBS in the dosage models but not the presence/absence models. We also identified statistically significant associations between specific MHC class II haplotypes and three fitness components: female LS, male ABS and male LBS. The associations between specific MHC haplotypes and female LS were identified in both dosage and presence/absence models while associations in male ABS and LBS were only identified in the dosage models.

As outlined above, evidence of MHC heterozygote fitness advantage has been reported in some previous studies (22, 23, 31) although other studies did not support this hypothesised mechanism. In addition, theoretical work has questioned the role of heterozygote advantage in maintaining MHC diversity and suggested that heterozygote advantage alone is not sufficient to account for the high degree of MHC polymorphism (46). In this study we only found a positive association between MHC heterozygosity and female JS in the presence/absence model which was not reflected in the dosage model. We did not find any positive associations between MHC heterozygosity and any adult fitness component. Indeed, we found that MHC heterozygosity was associated with decreased male LS and male LBS in the dosage model as well as decreased female LS in the presence/absence model which is contradictory to a fixed advantage for MHC heterozygosity. Overall, it seems unlikely that allelic diversity in MHC class II genes in Soay sheep is maintained by heterozygous advantage.

If heterozygote advantage does not explain the maintenance of MHC diversity then selection favouring or disfavouring particular haplotypes, as is likely under negative frequency-dependent selection when not in equilibrium or by the fluctuating selection hypothesis may be present (25, 35). Indeed, we found significant associations between specific MHC haplotypes and three adult fitness components. Haplotype D was associated with increased female life span while haplotypes F and H were associated with decreased female life span. Haplotype C was

281 associated with reduced male annual breeding success and lifetime breeding success. If these MHC-linked fitness effects are strong enough
282 and not in conflict with other effects, we should be able to see changes in haplotype frequency over time in response. Examination of effects
283 of MHC-fitness associations on both sexes suggests haplotype D often had positive coefficients in adult fitness components in both sexes, but
284 the patterns for the other haplotypes showing significant associations in specific fitness components are much less clear (Supplementary 4).
285 Consistent with this, the gene drop analysis showed that haplotype D has increased through the study period, although the change was not
286 significant for a two-tailed test (p=0.028). Taken together, modelling of fitness component associations and gene-drop simulations suggest
287 there is contemporary selection on MHC genes in Soay sheep, mainly comprising selection favouring haplotype D.
288
289 Regarding the specific mechanism of selection, one possibility is that the MHC class II genes in Soay sheep are under rare allele advantage
290 caused by an arms race between hosts and parasites. Previous theoretical work has suggested that frequency-dependent selection resulting
291 from host-parasite coevolution could favor novel alleles and is important in maintaining MHC polymorphism (47, 48). A recent empirical study
292 using guppies (*Poecilia reticulata*) and their common ectoparasite in controlled infections showed that novel MHC variants could reduce the
293 infection regardless of individual MHC variability and genetic background, which provides direct evidence of rare allele advantage in shaping
294 MHC diversity (49). Under rare allele advantage, fitness differences and frequency changes are only expected when the system is not in
295 equilibrium. If haplotypes are rarer than their equilibrium frequency, they will confer an advantage in fitness. Haplotype D is the rarest
296 haplotype in the Soay population with a very low haplotype frequency at the beginning of our study (only 7 individuals in the founder cohorts
297 1989-1992). Therefore, positive selection on haplotype D could come from rare-allele advantage (Figure 3).
298
299 However, our results are also consistent with the fluctuating selection hypothesis. Previous theoretical work showed that diversity of MHC
300 genes could be maintained by temporally fluctuating selection on its own in the absence of heterozygote and rare allele advantage if the
301 resistance alleles were dominant (15) although few empirical studies have found signatures of temporal fluctuation selection in a single
302 population. A study of the great reed warbler (*Acrocephalus arundinaceus*) identified significantly higher variation in MHC allele frequencies
303 between cohorts than at neutral loci suggesting that fluctuating selection may be acting on MHC genes (50). In our study, haplotype D may
304 confer the best resistance to the current parasite community on St Kilda under the fluctuating selection hypothesis.
305

306 Our results are thus consistent with both rare allele advantage and fluctuating selection and we are not capable of differentiating them. The
307 key differences between these two mechanisms are first, that fluctuating selection requires a changing external environment while rare allele
308 advantage is caused by an arms race between host and parasites, and second, that selection is directional for fluctuating selection while for
309 rare allele advantage selection is cyclical. However, not only are these differences hard to identify in real time, but also they are not mutually
310 exclusive, meaning that it is very hard to differentiate rare allele advantage and fluctuating selection in wild populations (12).
311
312 Several aspects of our analysis are novel in studies of MHC-fitness associations. To our knowledge, this is the first study which uses animal
313 models and includes a genomic inbreeding estimator in a study of association between MHC and fitness components in wild populations. As
314 the AM accounts for relatedness, we predicted that it would be more conservative and show fewer associations than GLMMs without
315 accounting for relatedness. In fact, we found the results to be largely consistent between the two models. This may be due to the generally
316 low heritability of the fitness traits (Supplementary 8). However, for traits which have a higher heritability, the results of animal models will be
317 more conservative as they would exclude the false positive effects caused by genomic regions other than the MHC. We are also the first to
318 examine MHC-fitness associations using both dosage models and presence/absence models in a wild population. We found that dosage
319 models and presence/absence models were generally consistent in terms of estimates (Figure 2). One possible explanation for this is that for
320 most haplotypes, homozygotes were relatively rare compared with heterozygotes. Thus, in some cases, there may not have been sufficient
321 power to differentiate dosage and presence/absence models.
322
323 Long term individual-based studies like the Soay sheep project also enable repeat measurements of the strength and direction of selection
324 over time (51, 52). In a previous study, three MHC-linked microsatellite alleles were found to be associated with survival in Soay sheep using
325 individuals from 1985 to 1994 (35). Microsatellite alleles OLADRB 257 and 205 were associated with decreased juvenile and yearling survival
326 respectively whereas the OLADRB 263 allele was associated with increased yearling survival. These microsatellite alleles can be linked to the
327 MHC class II haplotypes studied here: OLADRB allele 257 is linked to haplotype B (Supplementary 9). However in our analysis of juvenile
328 survival, we did not find any evidence for haplotype differences (Wald tests not significant; Table 2) although estimates for haplotype B in
329 juvenile survival are consistently negative in females. The inconsistency between the previous and current study is probably due to differences
330 in sample size and statistical methods. Additionally, as alleles 205 and 263 are associated with multiple haplotypes, and we have studied
331 annual survival rather than yearling survival, it is hard to draw further comparisons.

In summary, we used well-characterized MHC class II haplotypes to investigate MHC-linked fitness effects in Soay sheep. To our knowledge, our dataset, with more than 3000 individuals, is the largest yet used to study selection on MHC genes in a wild population. Our results support the existence of contemporary selection on MHC class II genes in Soay sheep and suggest rare allele advantage or fluctuating selection is the mechanism in shaping MHC diversity in Soay sheep rather than heterozygous advantage.

Acknowledgments

We thank the National Trust for Scotland for permission to work on St. Kilda and QinetiQ for logistics and support. We thank J. Pilkington, I. Stevenson and many volunteers who have helped with fieldwork on the island and all those who have contributed to keeping the project going over many years. W. Huang was supported by Edinburgh Global research scholarship. K. Ballingall received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 731014KB (VetBioNet) and acknowledges the support received from the Scottish Government's strategic research programme. The long-term project on St Kilda, the KASP genotyping and the SNP array genotyping were funded by the UK Natural Environment Research Council and the European Research Council.

References

1. Obbard DJ, Welch JJ, Kim KW, & Jiggins FM (2009) Quantifying Adaptive Evolution in the Drosophila Immune System. *Plos Genet* 5(10).
2. Acevedo-Whitehouse K & Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol Evol* 21(8):433-438.
3. 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(3):363-377.
4. Piertney SB & Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* 96(1):7-21.
5. Garrigan D & Hedrick PW (2003) Detecting adaptive molecular polymorphism, lessons from the MHC. *Am J Hum Genet* 73(5):375-375.
6. Apanius V, Penn D, Slev PR, Ruff LR, & Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 17(2):179-224.
7. Oppelt C, Starkloff A, Rausch P, Von Holst D, & Rodel HG (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. *Molecular ecology* 19(19):4155-4167.

357 8. Radwan J, Kuduk K, Levy E, LeBas N, & Babik W (2014) Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the
358 ornate dragon lizard. *Molecular ecology* 23(24):5966-5978.
359 9. Oliver MK, Telfer S, & Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in
360 the water vole (Arvicola terrestris). *Proceedings. Biological sciences* 276(1659):1119-1128.
361 10. Westerdahl H, Asghar M, Hasselquist D, & Bensch S (2012) Quantitative disease resistance: to better understand parasite-mediated selection on
362 major histocompatibility complex. *Proceedings. Biological sciences* 279(1728):577-584.
363 11. Jordan WC & Bruford MW (1998) New perspectives on mate choice and the MHC. *Heredity* 81:239-245.
364 12. Spurgin LG & Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *P R Soc B* 277(1684):979-988.
365 13. Takahata N & Nei M (1990) Allelic Genealogy under Overdominant and Frequency-Dependent Selection and Polymorphism of Major
366 Histocompatibility Complex Loci. *Genetics* 124(4):967-978.
367 14. Hughes AL & Nei M (1988) Pattern of Nucleotide Substitution at Major Histocompatibility Complex Class-I Loci Reveals Overdominant Selection.
368 *Nature* 335(6186):167-170.
369 15. Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution; international journal of organic evolution* 56(10):1902-1908.
370 16. Hill AVS*, et al.* (1991) Common West African Hla Antigens Are Associated with Protection from Severe Malaria. *Nature* 352(6336):595-600.
371 17. Rivero-de Aguilar J*, et al.* (2016) MHC-I provides both quantitative resistance and susceptibility to blood parasites in blue tits in the wild. *J Avian Biol*
372 47(5):669-677.
373 18. Westerdahl H, Stjernman M, Raberg L, Lannefors M, & Nilsson JA (2013) MHC-I Affects Infection Intensity but Not Infection Status with a Frequent
374 Avian Malaria Parasite in Blue Tits. *Plos One* 8(8).
375 19. Westerdahl H*, et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *P R Soc B* 272(1571):1511-1518.
376 20. Radwan J*, et al.* (2012) MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular ecology* 21(10):2469-2479.
377 21. Sutton JT, Nakagawa S, Robertson BC, & Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at
378 MHC immunity genes. *Mol Ecol* 20(21):4408-4420.
379 22. Dunn PO, Bollmer JL, Freeman-Gallant CR, & Whittingham LA (2013) Mhc Variation Is Related to a Sexually Selected Ornament, Survival, and
380 Parasite Resistance in Common Yellowthroats. *Evolution; international journal of organic evolution* 67(3):679-687.
381 23. Banks SC, Dubach J, Viggers KL, & Lindenmayer DB (2010) Adult survival and microsatellite diversity in possums: effects of major histocompatibility
382 complex-linked microsatellite diversity but not multilocus inbreeding estimators. *Oecologia* 162(2):359-370.
383 24. Lukasch B*, et al.* (2017) Major histocompatibility complex genes partly explain early survival in house sparrows. *Sci Rep-Uk* 7.
384 25. Sepil I, Lachish S, & Sheldon BC (2013) Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular ecology*
385 22(2):384-396.
386 26. Kloch A, Baran K, Buczek M, Konarzewski M, & Radwan J (2013) MHC influences infection with parasites and winter survival in the root vole
387 Microtus oeconomus. *Evol Ecol* 27(3):635-653.
388 27. Karlsson M*, et al.* (2015) House sparrowPasser domesticussurvival is not associated with MHC-I diversity, but possibly with specific MHC-I alleles.
389 *Journal of Avian Biology* 46(2):167-174.

390 28. Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Mol Ecol Resour* 10(2):237-251.
391 29. Babik W, Taberlet P, Ejsmond MJ, & Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC
392 system. *Mol Ecol Resour* 9(3):713-719.
393 30. Osborne AJ*, et al.* (2015) Heterozygote advantage at MHC DRB may influence response to infectious disease epizootics. *Molecular ecology*
394 24(7):1419-1432.
395 31. Brouwer L*, et al.* (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations.
396 *Molecular ecology* 19(16):3444-3455.
397 32. Clutton-Brock T & Pemberton JM (2004) Soay sheep - Dynamics and selection in an island population. *Cambridge University Press*.
398 33. Berenos C*, et al.* (2015) Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular ecology* 24(8):1810-1830.
399 34. Paterson S (1998) Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *J Hered* 89(4):289-294.
400 35. Paterson S, Wilson K, & Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in
401 a large unmanaged ungulate population (Ovis aries L.). *Proceedings of the National Academy of Sciences of the United States of America* 95(7):3714-
402 3719.
403 36. Charbonnel N & Pemberton J (2005) A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through
404 spatial and temporal fluctuations in selection. *Heredity* 95(5):377-388.
405 37. Dicks KL, Pemberton JM, & Ballingall KT (2019) Characterisation of major histocompatibility complex class IIa haplotypes in an island sheep
406 population. *Immunogenetics* 71(5-6):383-393.
407 38. Dicks K (2017) Unravelling major histocompatibility complex diversity in the Soay Sheep of St.Kilda. *University of Edinburgh*.
408 39. Berenos C, Ellis PA, Pilkington JG, & Pemberton JM (2014) Estimating quantitative genetic parameters in wild populations: a comparison of pedigree
409 and genomic approaches. *Mol Ecol* 23(14):3434-3451.
410 40. Huisman J (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*
411 17(5):1009-1024.
412 41. Morrissey MB*, et al.* (2012) The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to
413 the breeder's equation. *Evolution; international journal of organic evolution* 66(8):2399-2410.
414 42. Johnston SE, Berenos C, Slate J, & Pemberton JM (2016) Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a
415 Wild Population of Soay Sheep (Ovis aries). *Genetics* 203(1):583-+.
416 43. Berenos C, Ellis PA, Pilkington JG, & Pemberton JM (2016) Genomic analysis reveals depression due to both individual and maternal inbreeding in a
417 free-living mammal population. *Mol Ecol* 25(13):3152-3168.
418 44. Hadfield JD (2010) MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J Stat Softw* 33(2):1-22.
419 45. R Core Team (2013) R: A language and environment for statistical computing. .
420 46. De Boer RJ, Borghans JAM, van Boven M, Kesmir C, & Weissing FJ (2004) Heterozygote advantage fails to explain the high degree of polymorphism
421 of the MHC. *Immunogenetics* 55(11):725-731.

47. Ejsmond MJ & Radwan J (2015) Red Queen Processes Drive Positive Selection on Major Histocompatibility Complex (MHC) Genes. *Plos Comput Biol* 11(11).
48. Borghans JAM, Beltman JB, & De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55(11):732-739.
49. Phillips KP*, et al.* (2018) Immunogenetic novelty confers a selective advantage in host-pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America* 115(7):1552-1557.
50. Westerdahl H, Hansson B, Bensch S, & Hasselquist D (2004) Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *J Evol Biol* 17(3):485-492.
51. Siepielski AM, DiBattista JD, & Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecol Lett* 12(11):1261-1276.
52. Siepielski AM, DiBattista JD, Evans JA, & Carlson SM (2011) Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *P R Soc B* 278(1711):1572-1580.