An investigation of MHC-dependent sexual selection in a free-living population of sheep

Wei Huang, Jill G. Pilkington, Josephine M. Pemberton

Abstract

MHC genes are one of the most polymorphic gene clusters in vertebrates and play an essential role in adaptive immunity. There is evidence that sexual selection also plays a role in maintaining MHC diversity, but the specific mechanisms are controversial. In this study, we investigate evidence for non-random MHC-dependent mating patterns in a free-living population of Soay sheep. Using a large number of sheep diplotyped at the MHC class IIa region and genome-wide SNPs, together with field observations of consorts, we found sexual selection against one of eight haplotypes, C, in males at the pre-copulatory stage and sexual selection on female MHC heterozygosity acting in opposition directions at the pre- and post-copulatory stages. We also found disassortative mating at the post-copulatory stage, along with strong evidence of inbreeding avoidance at both stages. However, results from generalized linear mixed models suggest that the pattern of MHC-dependent disassortative mating could be a by-product of inbreeding avoidance. These results suggest that there is selection on the MHC at the pre- and post-copulatory stages, but that apparent disassortative mating with respect to the MHC may be driven by inbreeding avoidance.

Introduction

Major histocompatibility complex genes encode cell surface proteins that present pathogen-derived peptide to T cells to activate the adaptive immune response. It has been proposed that vertebrates may use MHC variation as a basis for sexual selection (Jordan and Bruford 1998, Penn 2002, Milinski 2006, Winternitz, Minchey et al. 2013, Ejsmond, Radwan et al. 2014, Radwan, Babik et al. 2020). Sexual selection via mate choice (intersexual selection) could increase or maintain genetic variation through both direct and indirect benefits (Neff and Pitcher 2005). To increase offspring fitness, an individual could choose a partner based on genetic quality (direct benefits) or based on genetic compatibility with themselves (indirect benefits). In addition, intrasexual selection through male-male competition could also contribute to genetic diversity, for example if the most competitive males are more heterozygous. In relation to intersexual selection, several studies have demonstrated that MHC genes could be a cue for kin recognition preventing mating with relatives through MHC-associated odours (Grob, Knapp et al. 1998, Penn 2002, Milinski 2006). However, there is a question whether this is a kin recognition system or whether it is a by-product of inbreeding avoidance due to a generalised ability to detect relatedness (Hurst, Thom et al. 2005, Sherborne, Thom et al. 2007).

MHC-dependent sexual selection could occur via three non-mutually exclusive processes: selection could favour MHC alleles or haplotypes, MHC diversity or MHC compatibility. First, individuals with specific MHC alleles could be favoured by intrasexual or intersexual selection and such a pattern has been reported in several studies (Ekblom, Saether et al. 2004, Eizaguirre, Yeates et al. 2009). In addition, some studies have reported that individuals with higher diversity or heterozygosity at the MHC were favoured as partners (Landry, Garant et al. 2001, Richardson, Komdeur et al. 2005, Cutrera, Fanjul et al. 2012, Winternitz, Promerova et al. 2015). In terms of MHC compatibility, MHC-dependent disassortative mating has been reported, which could maximise the MHC diversity of offspring (Schwensow, Eberle et al. 2008, Setchell, Charpentier et al. 2010, Hoover, Alcaide et al. 2018, Han, Sun et al. 2019). However, as excessive expression of MHC molecules could lead to depletion of the mature T-cell repertoire and elevated risk of autoimmune disease, it has also been suggested that individuals may optimise rather than maximise MHC diversity (Reusch, Haberli et al. 2001, Griggio, Biard et al. 2011, Rekdal, Anmarkrud et al. 2019). In addition, some studies have found mate choice favouring MHC-similar mates and explained such patterns as local adaption to endemic pathogens (Bonneaud, Chastel et al. 2006, Bichet, Penn et al. 2014, Sin, Annavi et al. 2015). Finally, some studies have found no evidence of MHC-dependent sexual selection and suggested it may not be a universal phenomenon (Paterson and Pemberton 1997, Westerdahl 2004, Huchard, Knapp et al. 2010, Sepil, Radersma et al. 2015, Yu, Nie et al. 2018). A meta-analysis of MHC-dependent sexual selection in non-model species found support for female choice for MHC diversity and selection for dissimilarity, but only when the MHC is characterized across multiple loci (Kamiya, O'Dwyer et al. 2014). Therefore, more empirical studies are needed to draw definite conclusions about MHC-dependent sexual selection.

Future studies on MHC-dependent sexual selection would ideally include advances in a number of aspects of data quality. First, accurate estimates of relatedness should be examined at the same time as MHC-dependent sexual selection. Rather than being the actual cues for sexual selection, MHC genes could be incidentally associated with signals of genome-wide relatedness. Thus, signatures of MHC-dependent disassortative mating could be a by-product of inbreeding avoidance. In the previous studies reported above (Bonneaud, Chastel et al. 2006, Setchell, Charpentier et al. 2010, Huchard, Baniel et al. 2013, Bichet, Penn et al. 2014, Winternitz, Promerova et al. 2015, Ferrandiz-Rovira, Allaine et al. 2016), relatedness was usually estimated from small panels of microsatellite markers rather than genome-wide SNPs, which may have been imprecise. Second, behavioural observations of mating are required to differentiate pre- and post-copulatory MHC-dependent sexual selection, especially since pre- and post-copulatory sexual selection on MHC genes could occur in opposite directions. For example, pre-copulatory sexual selection favouring MHC-dissimilar partners has been demonstrated in a population of salmon (*Salmo salar*) (Landry, Garant et al. 2001) while post-copulatory sexual selection favoured MHC-similar partners in another salmon study system (Yeates, Einum et al. 2009). However, few studies have examined whether MHC-dependent sexual selection occurs at the pre- and post-copulatory sexual selection stages using field observations of mating behaviour in the same population except a study of mouse lemur (Schwensow, Eberle et al. 2008). Finally, larger sample sizes are essential in future studies to secure confident results. A recent study documented the impact of sample size on error rates and effect sizes and suggested a sample size of 500 mating pairs is required for testing MHC-dependent sexual selection (Hoover and Nevitt 2016) which was not always available in previous studies.

In this study, we used a free-living population of Soay sheep (*Ovis aries*) on the island of Hirta, St Kilda to investigate MHC-dependent sexual selection. An individual-based study has been carried out on the population since 1985 (Clutton-Brock and Pemberton 2004). A large number of male-female consorts has been recorded during the rut each year and a multi-generation pedigree including most study individuals has been constructed. A previous study using five MHC-linked microsatellite loci genotyped between 887 and 1209 individuals born between 1985 and 1994 found no evidence for MHC-dependent assortative or disassortative mating in this population (Paterson and Pemberton 1997). Recently, using genotyping-by-sequencing, a total of eight MHC class II haplotypes have been identified in the study populations and a large number of individuals alive between 1989 and 2012 have been diplotyped (Dicks 2018, Dicks, Pemberton et al. 2019). In addition, pairwise relatedness based on genome-wide SNPs is available between most individuals. This genetic and genomic data combined with a large number of consort records and parentage records enabled us to test MHC-dependent sexual selection more thoroughly than before using Monto-Carlo simulations. We asked the following questions: 1) Are individuals carrying specific MHC haplotypes favoured during mating? 2) Are MHC-heterozygote individuals favoured during mating? 3) Is there MHC-dependent disassortative or assortative mating? 4) If there is disassortative mating, is it based on haplotype divergence? 5) Is there inbreeding preference or avoidance and if so how is related to any MHC-based mating

patterns? 6) If any signature of non-random mating is detected, does it occur at the pre- or post-copulatory stage?

Methods

Study population and data collection

An unmanaged population of Soay sheep has resided unmanaged on the island of Hirta, St Kilda since 1932 when they were introduced there from the nearby island of Soay. From 1985, a longitudinal individual-based study has been conducted on the sheep resident in the Village Bay area of Hirta. Nearly all individual Soay sheep living in the study area have been followed from birth, through all breeding attempts, to death. Lambs born as singletons, twins or (rarely) triplets are ear tagged shortly after birth in April or May, sampled for genetic analysis and weighed. Any missed lambs or immigrant adults are captured, tagged and sampled in an August catch up or in the rut in November. The population is regularly censused throughout the year with individual locations recorded (Clutton-Brock and Pemberton 2004).

A large number of study area sheep alive since 1989 have been genotyped on the Illumina Ovine 50K SNP array. Parentage was determined for each individual using a subset of 315 unlinked SNPs derived from the SNP array using the R library Sequoia (Berenos, Ellis et al. 2014, Huisman 2017).

Rut behavioural data

Soay sheep have a promiscuous mating system with intensive male-male competition as well as male mate choice (Preston, Stevenson et al. 2003, Preston, Stevenson et al. 2005). The onset of the rut in early November is marked by increasing male aggression as males roam and search for oestrous females across the study area. Males compete to gain access to oestrous females and the winner defends and mates with the female repeatedly over several hours in a so-called 'consort'. However, only large and mature males with big horns can defend a female for long. Younger, smaller males and those with scurred horns constantly search for oestrous females, chase them and get quick matings if they can. Matings are mostly between males and females aged one year or older, but male lambs aged seven months obtain matings and some female lambs give birth at the age of one (see Supplementary table 1). Until the end of the rut around early December, the study area is continually monitored for consorts throughout each day, with consorts defined as being a close spatial relationship between a male and female with frequent male courtship and defence of a receptive female (Clutton-Brock and Pemberton 2004).

Molecular data

The molecular data used in this study included MHC class II diplotypes and pairwise genomic relatedness. We used a two-step haplotyping method involving characterisation of the MHC haplotypes present and then Kompetitive Allele-specific PCR (KASP) genotyping to impute haplotypes of individuals that lived in the study area between 1989 and 2012. First, seven

157 expressed loci (DRB1, DQA1, DQA2, DQA2-like, DQB1, DQB2 and DQB2-like) within the MHC
158 class IIa region were characterized in 118 Soay sheep using genotyping-by-sequencing which
159 identified eight haplotypes named A to H (Dicks, Pemberton et al. 2019). Second, a panel of
160 13 SNPs located in the MHC class IIa region haplotypes including 11 SNPs from the Ovine
161 Infinium HD chip and two other SNPs located within DQA1 gene were selected to impute
162 eight haplotypes using Kompetitive Allele-specific PCR (KASP) in 5951 Soay sheep (Dicks
163 2018, Dicks, Pemberton et al. 2020). After genotyping quality control and pedigree checking,
164 5349 individuals were successfully diplotyped. The pairwise genomic relatedness between
165 all individuals was calculated using DISSECT (Canela-Xandri, Law et al. 2015) based on 37K
166 polymorphic SNPs from the Illumina Ovine 50K SNP array excluding the X chromosome and
167 chromosome 20 where the MHC genes are located.

168 Assortative mating analysis

169 We performed Monto-Carlo simulations to examine whether there were MHC-dependent or
170 relatedness-dependent mating patterns in Soay sheep.

171 We first selected all females and males older than one year which were diplotyped at the
172 MHC and genotyped on the SNP chip into a "primary mating pool". We focused on mating
173 between adult sheep because relatively few offspring have a juvenile parent (that is, a male
174 lamb for a father and/or a female lamb for a mother; see Supplementary table 1) and
175 including the large number of non-reproductive male and female lambs present each
176 breeding season could have biased the simulations. We excluded individuals whose birth
177 year was unknown, as they do not live mainly in our study area and we had no information
178 about when they start to be involved in mating. For the few individuals with no recorded
179 death year, death year was estimated from the last-seen year recorded in the census data
180 or the last year they sired or gave birth to offspring. A total of 889 females and 821 males
181 were included in the primary mating pool dataset and each of them is recorded once per
182 year they were alive (Supplementary table 1).

183 Second, we extracted all the consorts between any male and female in the primary mating
184 pool dataset as the "consort dataset". Multiple observations of the same pair together on
185 the same day were counted as one consort observation.

186 Third, we assembled an "observed parentage dataset" comprising all mother-father-
187 offspring trios in which the offspring birth year was known, all three individuals were
188 successfully diplotyped at the MHC and the parents were genotyped on the SNP chip. To be
189 consistent between the primary mating pool and observed parentage dataset, we excluded
190 trios in which either of the parents was a lamb or an adult not included in primary mating
191 pool, such that all parents could be sampled from the primary mating pool. In total, 2068
192 trios were included in the observed parentage dataset (Supplementary table 2). Twins and
193 triplets were not common in our dataset comprising less than a quarter of all individuals
194 born. Twins are usually half-sibs with different fathers, coming from different mating events
195 (Supplementary table 3). Thus, we treated each offspring as an independent data point.

196 Finally, three null models of random mating were defined for simulation.

1) Primary random mating model:

Each offspring in the observed parentage dataset was randomly assigned a father and mother in the year before its birth year from the primary mating pool. Most female sheep have a single lamb per year, but they can have up to 3, while male sheep have a great range of annual breeding success. Thus, we sampled fathers with replacement but mothers without replacement. As expected from this approach, annual and lifetime breeding success of both sexes had a smaller range than the observed parentage dataset (Supplementary figure 1).

2) Adjusted random mating model:

For each year, we first randomly assigned a male or a female living in the same year from the primary mating pool to replace each known mother and father in the observed parentage dataset and then adjusted the record of the sampled sheep based on the annual breeding success of the replaced sheep to produce an "adjusted mating pool". Then, each offspring in the observed parentage dataset was randomly assigned a father and mother in the year before its birth year without replacement from the adjusted mating pool. Annual breeding success in the simulated results was the same as the observed parentage dataset, while lifetime breeding success differed slightly from the observed parentage dataset with the mean of lifetime breeding success lower in the simulated data than that from the observed parentage dataset (Supplementary figure 2).

3) Restricted random mating model:

For each year, we selected adult males and females that had offspring in the following year into a "restricted mating pool" so their records corresponded with the number of offspring they had. Each offspring in the "observed parentage dataset" was randomly assigned a father and mother living in the year before its birth year without replacement from the restricted mating pool. As expected, both annual and lifetime breeding success in the simulated results were identical to the observed parentage dataset (Supplementary figure 3).

All three models were simulated for 10000 iterations in R v.3.5.2 using a custom script (R Core Team 2013).

We believe that the adjusted random mating model is the most appropriate model of the mating system as it captures the intense male-male competition in our system by accounting for annual breeding success while it also allows for a wide range of individuals in the mating pool rather than only simulating individuals known to be parents. Nevertheless, we also used the primary and restricted random models to capture more random and more restricted scenarios. Previous studies have suggested accounting for spatial distance when designing null models of random mating (Huchard, Baniel et al. 2013, Sepil, Radersma et al. 2015). However, during the rut, male sheep rove around the entire study area to search for oestrous females such that there is little evidence of spatially-restricted mating patterns (Clutton-Brock and Pemberton 2004). Therefore, we did not account for spatial distance in our null models.

After simulation, we summarised the results of all the iterations using various indices in response to the questions we proposed: 1) The frequency of each haplotype in simulated mothers and fathers. 2) The ratio of homozygote: heterozygote in simulated mothers and fathers. 3) The average number of shared MHC haplotypes between simulated parents and the proportion of simulated parents sharing 0, 1 and 2 haplotypes. 4) To account for MHC functional variation, the pairwise divergence of MHC haplotypes between each simulated parent pair was first measured by the proportion of the amino acid sequence that differed between them (p-distance) (Henikoff 1996). We then defined two indices AAdist and distmax as the mean and maximum MHC divergence between the 4 possible haplotype combinations of each simulated parent pair respectively. Finally, the mean of AAdist and distmax were calculated. 5) The mean and median of genomic relatedness between simulated parents.

These indices were also calculated for the real data using the "consort dataset" and "observed parentage dataset". Specifically, the first two indices were measured in mated females and males (consort dataset) or in mothers and fathers (observed parentage dataset) while the last three indices were measured between consort pairs or between genetic parents. For each index, statistical significance (p value) was determined by comparing the index in the real data with the 2.5% and 97.5% tails of the distribution of the index in the simulated results. To account for multiple testing, we applied a Bonferroni correction to the eight haplotype frequency tests across two sexes, and significant results were determined based on the refined critical p-value ($p = 0.0015625$).

Once significant results were identified, we determined whether the distortion occurred at the pre- or post-copulatory stage based on the following logic. 1) If indices in the consort dataset and observed parentage dataset both deviate from expectations of random mating, this is evidence of pre-copulatory selection. 2) If indices in the consort dataset and observed parentage dataset both deviate from expectations of random mating but in opposite directions, or only indices in the consort dataset deviated, this is evidence of both pre- and post-copulatory selection acting in different directions. 3) If only indices in the observed parentage dataset deviate from expectations of random mating, this is evidence of post-copulatory selection. 4) If indices in the consort dataset and observed parentage dataset were both in line with expectations of random mating, this suggests no evidence of sexual selection.

Generalized linear mixed models

We used generalized linear mixed models to differentiate genomic inbreeding avoidance and MHC-dependent disassortative mating in Soay sheep. We drew up a matrix of all pairwise combinations of males and females in the primary mating pool in a given year. For each pair, we then recorded their consort and breeding success (0/1) based on the real data from consort dataset and observed parentage dataset respectively, with 0 meaning no successful consort or offspring observed and 1 meaning consort or offspring observed. Then, we investigated consort and breeding success as response in separate binomial regressions. In each model, year and sire ID were fitted as random effects while genomic

278 relatedness and the number of shared MHC haplotypes between each pair were fitted
279 simultaneously as fixed effects. The model was run in R v.3.5.2 using R package lme4.

280

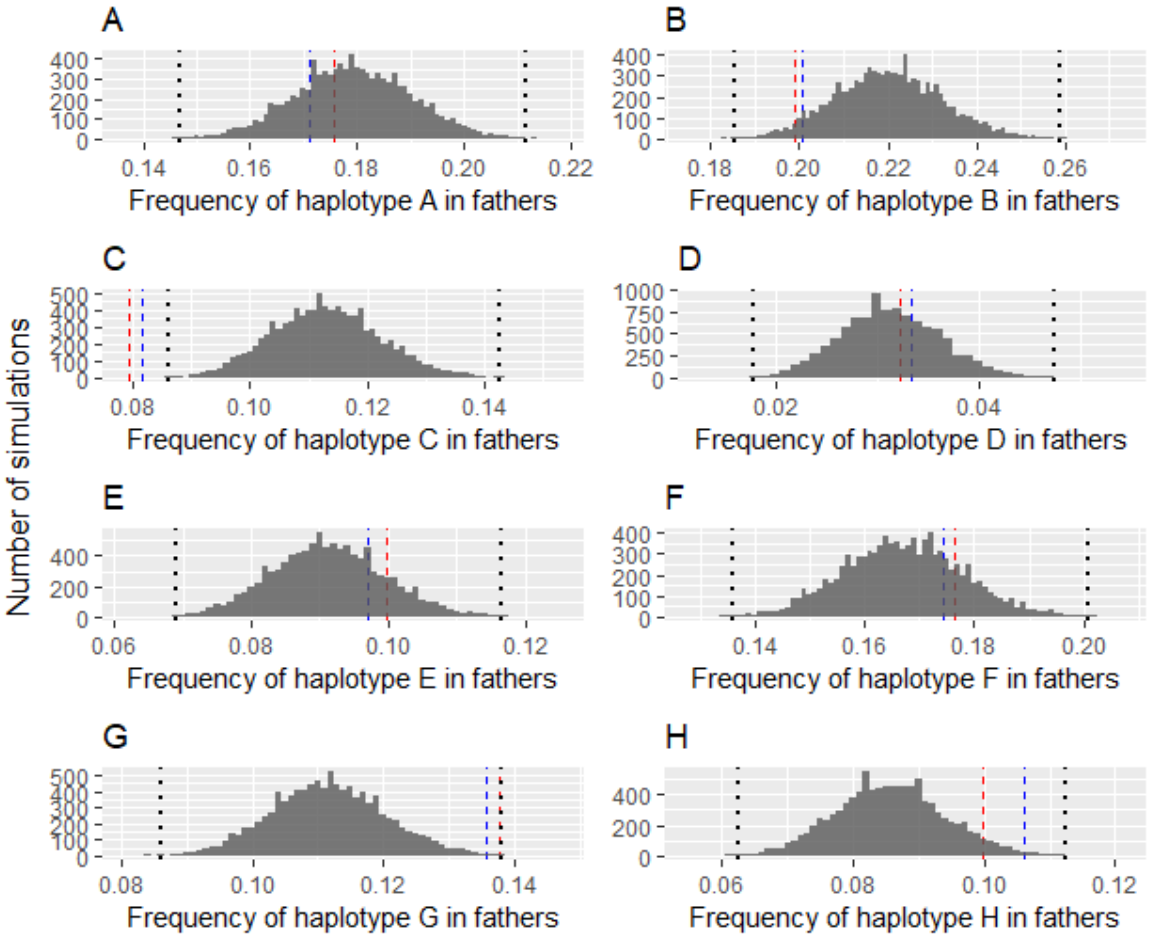
Results

We used three null models to examine MHC-dependent sexual selection. Here we focus on the results from the adjusted random mating model (Figures 1-3, Supplementary Figure 9). The results of the primary and the restricted random mating models are described thoroughly in the supplementary information (Supplementary Figures 4-8).

Haplotype frequency tests

The frequency of haplotype C in both consort males and fathers was significantly lower than expected under the null model. In addition, the frequency of haplotype G in both consort males and fathers and the frequency of haplotype H in fathers tended to be higher than expected but these results were not significant after Bonferroni correction (Figure 1).

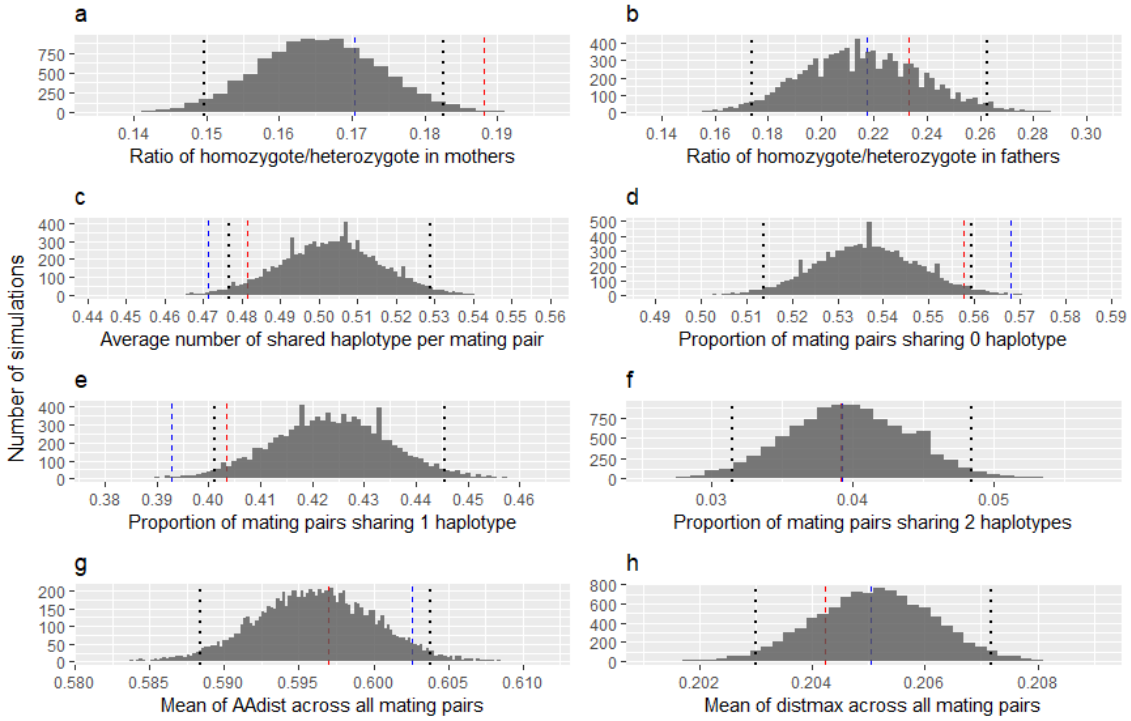
We did not find any significant deviation of haplotype frequency in either consort females or mothers relative to random mating. The frequency of haplotype F and H in consort females tended to be lower and higher respectively. In addition, the frequency of haplotypes A and G in mothers tended to be lower and higher respectively. However, none of these patterns was significant after Bonferroni correction (Supplementary figure 4).



297 Figure 1. Results of MHC haplotype frequency tests in fathers under the adjusted random
298 mating model. Histograms represent the result of simulations with dotted black lines
299 representing the critical p-values after Bonferroni correction. The red and blue dashed blue
300 lines show the observed MHC haplotype frequency in the consort and observed parentage
301 dataset respectively.

302 **Diplotype-based tests**

303 Regarding individual MHC heterozygosity, we found that the ratio of homozygote to
304 heterozygote in consort females was significantly higher than expected under the null
305 model However, the ratio in mothers was in line with expectation under the null model
306 (Figure 2a). We found the average number of shared haplotypes between parents was
307 significantly lower compared with the null expectation but this pattern was not observed in
308 consort dataset (Figure 2c). In addition, the proportion of parents sharing 0 haplotype was
309 significantly higher than in the simulated results while the proportion of parents sharing 1
310 haplotype was significantly lower than the simulated results (Figure 2d-f). However, MHC
311 divergence measured as amino acid sequence differences between parents was in line with
312 expectation of random mating (Figure 2g and h).
313

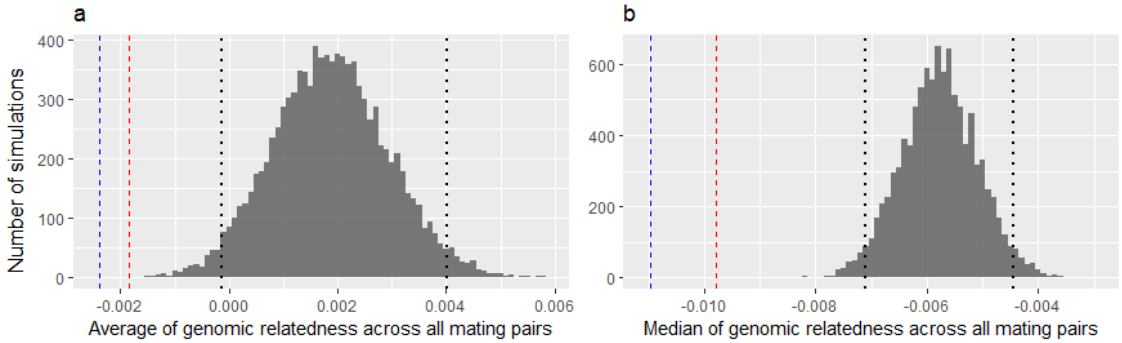


314

315 Figure 2. Results of MHC diplotype-based tests under the adjusted random mating model.
316 Histograms represent the result of simulations, with dotted lines representing the 2.5% and
317 97.5% tails of the distributions. The red and blue dashed blue lines show observed values for
318 the indices in the consort and observed parentage datasets respectively.

319 **Genome-wide relatedness tests**

320 We found mean and median genomic relatedness between consort pairs and those
321 between parents were significantly lower than expected under the null model (Figure 3).



322

323 Figure 3. Results of tests of pairwise genomic relatedness under the adjusted random
324 mating model. Histograms represent the result of simulations, with dotted lines
325 representing the 2.5% and 97.5% tails of the distributions. The red and blue dashed blue
326 lines show observed values for genomic relatedness in the consort and observed parentage
327 dataset respectively.

328 **Results from the primary and restricted random mating models**

329 We found there were more significant results in haplotype frequency tests under the
330 primary random mating model than those in the adjusted random mating model
331 (Supplementary figure 5 and 6). Nevertheless, the haplotype C result derived from the
332 adjusted random mating model was also detected in the primary mating pool model
333 (Supplementary figure 5). However, our results were generally consistent between the
334 primary and adjusted mating models for the diplotype-based tests and genome-wide
335 relatedness tests (Supplementary figure 7 and 8). The signature of sexual selection on MHC
336 heterozygosity in mothers and that on MHC compatibility were recovered under the primary
337 random mating model. In addition, signature of genomic inbreeding avoidance was
338 identified.

339 Fewer significant results were detected under the restricted mating model (Supplementary
340 figure 9). First, evidence of sexual selection on MHC compatibility was not fully detected.
341 Under the restricted mating model, we only found the simulated median of genomic
342 relatedness were lower than those in the observed parentage dataset but not in the consort
343 dataset. However, we found average MHC divergence between parents were significantly
344 higher than the simulated results under restricted random mating model which was not
345 observed in the primary and adjusted mating model.

346 **Generalized linear mixed models**

347 We found a significant association between both consort and breeding success and genomic
348 relatedness, but no significant association between consort and breeding success and
349 number of shared MHC haplotypes (Table 1).

350

Table 1. Results of Generalized linear mixed model testing associations between consort/breeding success and MHC compatibility/genomic relatedness. Significant effects are marked with asterisk (* $p<0.05$, ** $p<0.01$)

	Consort Success	Breeding success
Fixed effects		
Number of shared MHC haplotypes	-0.029 (0.031)	-0.067(0.041)
Genomic relatedness	-1.106 (0.362)**	-0.947(0.477)*
Random effects		
Sire ID	3.917(1.979)	1.844(1.358)
Year	1.276(1.130)	0.362(0.602)

Discussion

In this study, we investigated MHC-dependent sexual selection in a free-living sheep population using Monto-Carlo simulations. By comparing the result of simulated and real data using three null models, we examined whether there is deviation from random mating depending on MHC genes. The adjusted random mating model was considered the best model, so we mainly consider the results derived from this model in the following discussion. We found haplotype C in both consort males and fathers was disfavoured in comparison with random expectation. We found no evidence that MHC heterozygotes were favoured in either sex. Instead, we found MHC homozygote females were over-represented in consort pairs but this pattern was not observed in actual mothers. We found the average number of shared haplotypes between parents was lower compared with null expectation, but this pattern was not observed in the consort dataset. In addition, the proportion of parent pairs sharing no haplotype was higher and that of parent pairs sharing one haplotype was lower than expected under random mating. Finally, we found evidence of inbreeding avoidance, as the mean and median pairwise genomic relatedness in the consort and observed parentage datasets were significantly lower than expected under random mating. When fitting MHC haplotype sharing and genomic relatedness in the same model of consort or parentage, we could not demonstrate an independent effect of disassortative mating based on MHC haplotype sharing.

In our study, sexual selection on a specific MHC haplotype (C) was probably due to differences in male competitive ability, since the frequency of C was rarer than expected, not only in male parents but also in consort males. Few previous studies have reported specific MHC variants being favoured or disfavoured during mating as the high polymorphism of MHC genes requires a large sample size to detect sexual selection on specific MHC variants (Eizaguirre, Yeates et al. 2009). In this study, our finding for haplotype C in male parents was consistent with a negative association between MHC haplotype C and male breeding success found in a recent study on MHC-fitness associations (Huang, Dicks et al. 2020). The fact that haplotype C males are also less often observed in consort

384 than expected indicates that the effect of haplotype C is expressed at the pre-copulatory
385 rather than post-copulatory stage.

386 Our finding that females involved in consorts are more homozygous than expected, an
387 effect which is not found in the observed parentage data, is puzzling. One possibility is that
388 MHC-homozygous females are less likely to conceive in a given oestrus and therefore cycle
389 and return to oestrus 14 days later. This is turn would enrich our consort data set for such
390 females. Alternatively, if homozygous females are less attractive in some way, perhaps they
391 are less likely to be in long, stable consorts and instead experience multiple short consorts,
392 which would again enrich the consort data set for homozygous females. These possibilities
393 require further investigation within our dataset but are beyond the scope of this paper.

394 At first sight, our results also suggest that there is sexual selection based on MHC
395 compatibility, but our tests suggest this effect is not independent of an inbreeding
396 avoidance effect. In a population with limited dispersal and severe inbreeding depression,
397 inbreeding avoidance could arise to reduce the cost of inbreeding through kin recognition
398 (Pusey and Wolf 1996, Szulkin, Stopher et al. 2013, Duthie and Reid 2016). Previous studies
399 have proposed MHC genes could be the cues for inbreeding avoidance and MHC-associated
400 odour has been reported in a wide range of taxa including fish (Olsen, Grahn et al. 1998,
401 Milinski, Griffiths et al. 2005), reptiles (Olsson, Madsen et al. 2003), birds (Leclaire, Strandh
402 et al. 2017) and mammals (Wedekind, Seebeck et al. 1995, Wedekind and Furi 1997,
403 Yamazaki, Beauchamp et al. 2000, Roberts, Gosling et al. 2008). However, if MHC haplotype
404 sharing is associated with relatedness, as we have shown in Soays, MHC-disassortative
405 mating could be a by-product of inbreeding avoidance. Few studies have been able to test
406 this hypothesis, however a study of grey mouse lemur (*Microcebus murinus*) revealed both
407 inbreeding avoidance and MHC-dependent disassortative mating, and suggested that
408 observed deviations from random mating at the MHC are driven by the functionally
409 important MHC gene DRB rather than resulting passively from inbreeding avoidance. In that
410 study, MHC-dependent disassortative mating was detected at the DRB locus only for amino
411 acid sequence and functional similarity rather than number of shared MHC alleles (Huchard,
412 Baniel et al. 2013). In contrast, when studied in terms of MHC divergence measured as
413 amino acid sequence differences between parents in Soay sheep, our results were in line
414 with the expectation from random mating (Supplementary Figure 11). In Soay sheep,
415 inbreeding depression has been documented repeatedly using different markers and
416 methods (Coltman, Pilkington et al. 1999, Overall, Byrne et al. 2005, Berenos, Ellis et al.
417 2016, Stoffel, Johnston et al. 2020) such that inbreeding avoidance could be occuring in
418 response to inbreeding depression. Using genomic relatedness calculated from a large
419 number of SNPs, we demonstrated genomic inbreeding avoidance in Soays at the pre-
420 copulatory stage for the first time. When tested in the same model, MHC haplotype sharing
421 was not significant. We therefore cannot claim that the apparent disassortative mating
422 based on haplotype sharing is anything but a correlated effect of inbreeding avoidance.

423 Our results differ from a previous study in Soay sheep (Paterson and Pemberton 1997)
424 which found no evidence for MHC-dependent assortative or disassortative mating. In the
425 current study, we found evidence of MHC-dependent disassortative mating, at the post-

426 copulatory stage and which carried through to the parentage stage, based on the number of
427 shared MHC haplotypes. Reasons for this difference include the fact our data consisted MHC
428 class II haplotypes rather than MHC-linked microsatellite markers, and these two
429 approaches do not have a perfect read through. Also, our sample sizes are very much larger
430 and our methodology (Monto-Carlo simulation) is different from the previous study
431 (likelihood-based approach).

432 By using a large number of consort observations, we were able to differentiate MHC-
433 dependent sexual selection at the pre- and post-copulatory stages. In this area, field
434 observations were first used in a study of mouse lemur which demonstrated post-copulatary
435 MHC-dependent disassortative mating (Schwensow, Eberle et al. 2008). Here, we used both
436 the consort and the observed parentage datasets to examine MHC-dependent sexual
437 selection. We found sexual selection against a specific MHC haplotype at pre-copulatory
438 stage and sexual selection favouring MHC compatibility at the post-copulatory stage. More
439 interestingly, we found that the ratio of homozygote:heterozygote was significantly higher
440 in consort females than the simulated results but this pattern were not observed in
441 mothers. These results suggested the importance of using field observations to differentiate
442 pre- and post-copulatory sexual selection.

443 In this study, we examined whether there was MHC-dependent sexual selection in a
444 population of free-living Soay sheep. Benefiting from intermediate MHC polymorphism, high
445 quality genetic and genomic information, intensive field observations and large sample size,
446 we have demonstrated sexual selection based on a specific MHC haplotype at the pre-
447 copulatory stage and MHC compatibility at the post-copulatory stage occurs simultaneously.
448 We have also demonstrated sexual selection on female MHC heterozygosity at both pre-
449 and post-copulatory stages, but the direction of selection was antagonistic. Finally, we
450 report inbreeding avoidance in this population for the first time and find that because
451 haplotype sharing is correlated with relatedness we cannot show an independent effect of
452 disassortative mating based on the MHC.

453 # Acknowledgements

454 We thank the National Trust for Scotland for permission to work on St. Kilda and QinetiQ,
455 Eurest and Kilda Cruises for logistics and support. We thank J. Pilkington, I. Stevenson and
456 many volunteers who have collected field data and samples, especially in the rut, and all
457 those who have contributed to keeping the project going. The MHC diplotyping method and
458 the diplotype data set were developed and generated by Kara Dicks and Susan Johnston.
459 SNP genotyping was conducted at the Wellcome Trust Clinical Research Facility Genetics
460 Core. Field data collection has been supported by NERC over many years, the diplotyping
461 was supported by the BBSRC and Royal Society and most of the SNP genotyping was
462 supported by the ERC. Wei Huang is supported by FILL IN NAME OF YOUR STUDENTSHIP

463

References

Berenos, C., P. A. Ellis, J. G. Pilkington and J. M. Pemberton (2014). "Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches." Molecular Ecology **23**(14): 3434-3451.

Berenos, C., P. A. Ellis, J. G. Pilkington and J. M. Pemberton (2016). "Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population." Molecular Ecology **25**(13): 3152-3168.

Bichet, C., D. J. Penn, Y. Moodley, L. Dunoyer, E. Cellier-Holzem, M. Belvalette, A. Gregoire, S. Garnier and G. Sorci (2014). "Females tend to prefer genetically similar mates in an island population of house sparrows." Bmc Evolutionary Biology **14**.

Bonneaud, C., O. Chastel, P. Federici, H. Westerdahl and G. Sorci (2006). "Complex Mhc-based mate choice in a wild passerine." Proceedings of the Royal Society B-Biological Sciences **273**(1590): 1111-1116.

Canela-Xandri, O., A. Law, A. Gray, J. A. Woolliams and A. Tenesa (2015). "A new tool called DISSECT for analysing large genomic data sets using a Big Data approach." Nature Communications **6**.

Clutton-Brock, T. H. and J. M. Pemberton (2004). Soay sheep: dynamics and selection in an island population, Cambridge University Press.

Coltman, D. W., J. G. Pilkington, J. A. Smith and J. M. Pemberton (1999). "Parasite-mediated selection against inbred Soay sheep in a free-living, island population." Evolution **53**(4): 1259-1267.

Cutrera, A. P., M. S. Fanjul and R. R. Zenuto (2012). "Females prefer good genes: MHC-associated mate choice in wild and captive tuco-tucos." Animal Behaviour **83**(3): 847-856.

Dicks, K. (2018). Unravelling major histocompatibility complex diversity in the Soay sheep of St Kilda.

Dicks, K. L., J. M. Pemberton and K. T. Ballingall (2019). "Characterisation of major histocompatibility complex class IIa haplotypes in an island sheep population." Immunogenetics **71**(5-6): 383-393.

Dicks, K. L., J. M. Pemberton, K. T. Ballingall and S. E. Johnston (2020). "Haplotyping MHC class IIa by high throughput screening in an isolated sheep population." bioRxiv: 2020.2007.2020.212225.

Duthie, A. B. and J. M. Reid (2016). "Evolution of Inbreeding Avoidance and Inbreeding Preference through Mate Choice among Interacting Relatives." American Naturalist **188**(6): 651-667.

Eizaguirre, C., S. E. Yeates, T. L. Lenz, M. Kalbe and M. Milinski (2009). "MHC-based mate choice combines good genes and maintenance of MHC polymorphism." Molecular Ecology **18**(15): 3316-3329.

Ejsmond, M. J., J. Radwan and A. B. Wilson (2014). "Sexual selection and the evolutionary dynamics of the major histocompatibility complex." Proceedings of the Royal Society B-Biological Sciences **281**(1796).

Ekblom, R., S. A. Saether, M. Grahn, P. Fiske, J. A. Kalas and J. Hoglund (2004). "Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (Gallinago media)." Molecular Ecology **13**(12): 3821-3828.

Ferrandiz-Rovira, M., D. Allaine, M. P. Callait-Cardinal and A. Cohas (2016). "Mate choice for neutral and MHC genetic characteristics in Alpine marmots: different targets in different contexts?" Ecology and Evolution **6**(13): 4243-4257.

Griggio, M., C. Biard, D. J. Penn and H. Hoi (2011). "Female house sparrows "count on" male genes: experimental evidence for MHC-dependent mate preference in birds." Bmc Evolutionary Biology **11**.

Grob, B., L. A. Knapp, R. D. Martin and G. Anzenberger (1998). "The major histocompatibility complex and mate choice: Inbreeding avoidance and selection of good genes." Experimental and Clinical Immunogenetics **15**(3): 119-129.

Han, Q. H., R. N. Sun, H. Q. Yang, Z. W. Wang, Q. H. Wan and S. G. Fang (2019). "MHC class I diversity predicts non-random mating in Chinese alligators (Alligator sinensis)." Heredity **122**(6): 809-818.

Henikoff, S. (1996). "Scores for sequence searches and alignments." Current Opinion in Structural Biology **6**(3): 353-360.

513 Hoover, B., M. Alcaide, S. Jennings, S. Y. W. Sin, S. V. Edwards and G. A. Nevitt (2018). "Ecology can
514 inform genetics: Disassortative mating contributes to MHC polymorphism in Leach's storm-petrels
515 (Oceanodroma leucorhoa)." Mol Ecol.
516 Hoover, B. and G. Nevitt (2016). "Modeling the Importance of Sample Size in Relation to Error in
517 MHC-Based Mate-Choice Studies on Natural Populations." Integrative and Comparative Biology
518 **56**(5): 925-933.
519 Huang, W., K. L. Dicks, J. D. Hadfield, S. E. Johnston, K. T. Ballingall and J. M. Pemberton (2020). "A
520 rare MHC haplotype confers selective advantage in a free-living ruminant." bioRxiv:
521 2020.2003.2025.008565.
522 Huchard, E., A. Baniel, S. Schliehe-Diecks and P. M. Kappeler (2013). "MHC-disassortative mate
523 choice and inbreeding avoidance in a solitary primate." Molecular Ecology **22**(15): 4071-4086.
524 Huchard, E., L. A. Knapp, J. Wang, M. Raymond and G. Cowlishaw (2010). "MHC, mate choice and
525 heterozygote advantage in a wild social primate." Mol Ecol **19**(12): 2545-2561.
526 Huisman, J. (2017). "Pedigree reconstruction from SNP data: parentage assignment, sibship
527 clustering and beyond." Molecular Ecology Resources **17**(5): 1009-1024.
528 Hurst, J. L., M. D. Thom, C. M. Nevison, R. E. Humphries and R. J. Beynon (2005). "MHC odours are
529 not required or sufficient for recognition of individual scent owners." Proceedings of the Royal
530 Society B-Biological Sciences **272**(1564): 715-724.
531 Jordan, W. C. and M. W. Bruford (1998). "New perspectives on mate choice and the MHC." Heredity
532 **81**: 239-245.
533 Kamiya, T., K. O'Dwyer, H. Westerdahl, A. Senior and S. Nakagawa (2014). "A quantitative review of
534 MHC-based mating preference: the role of diversity and dissimilarity." Molecular Ecology **23**(21):
535 5151-5163.
536 Landry, C., D. Garant, P. Duchesne and L. Bernatchez (2001). "'Good genes as heterozygosity': the
537 major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar)." Proceedings of
538 the Royal Society B-Biological Sciences **268**(1473): 1279-1285.
539 Leclaire, S., M. Strandh, J. Mardon, H. Westerdahl and F. Bonadonna (2017). "Odour-based
540 discrimination of similarity at the major histocompatibility complex in birds." Proceedings of the
541 Royal Society B-Biological Sciences **284**(1846).
542 Milinski, M. (2006). "The major histocompatibility complex, sexual selection, and mate choice."
543 Annual Review of Ecology Evolution and Systematics **37**: 159-186.
544 Milinski, M., S. Griffiths, K. M. Wegner, T. B. H. Reusch, A. Haas-Assenbaum and T. Boehm (2005).
545 "Mate choice decisions of stickleback females predictably modified by MHC peptide ligands."
546 Proceedings of the National Academy of Sciences of the United States of America **102**(12): 4414-
547 4418.
548 Neff, B. D. and T. E. Pitcher (2005). "Genetic quality and sexual selection: an integrated framework
549 for good genes and compatible genes." Mol Ecol **14**(1): 19-38.
550 Olsen, K. H., M. Grahn, J. Lohm and A. Langefors (1998). "MHC and kin discrimination in juvenile
551 Arctic charr, Salvelinus alpinus (L.)." Animal Behaviour **56**: 319-327.
552 Olsson, M., T. Madsen, J. Nordby, E. Wapstra, B. Ujvari and H. Wittsell (2003). "Major
553 histocompatibility complex and mate choice in sand lizards." Proceedings of the Royal Society B-
554 Biological Sciences **270**: S254-S256.
555 Overall, A. D. J., K. A. Byrne, J. G. Pilkington and J. M. Pemberton (2005). "Heterozygosity, inbreeding
556 and neonatal traits in Soay sheep on St Kilda." Molecular Ecology **14**(11): 3383-3393.
557 Paterson, S. and J. M. Pemberton (1997). "No evidence for major histocompatibility complex-
558 dependent mating patterns in a free-living ruminant population." Proceedings of the Royal Society B-
559 Biological Sciences **264**(1389): 1813-1819.
560 Penn, D. J. (2002). "The scent of genetic compatibility: Sexual selection and the major
561 histocompatibility complex." Ethology **108**(1): 1-21.

562 Preston, B. T., I. R. Stevenson, J. M. Pemberton, D. W. Coltman and K. Wilson (2003). "Overt and
563 covert competition in a promiscuous mammal: the importance of weaponry and testes size to male
564 reproductive success." Proceedings of the Royal Society B-Biological Sciences **270**(1515): 633-640.
565 Preston, B. T., I. R. Stevenson, J. M. Pemberton, D. W. Coltman and K. Wilson (2005). "Male mate
566 choice influences female promiscuity in Soay sheep." Proceedings of the Royal Society B-Biological
567 Sciences **272**(1561): 365-373.
568 Pusey, A. and M. Wolf (1996). "Inbreeding avoidance in animals." Trends in Ecology & Evolution
569 **11**(5): 201-206.
570 R Core Team (2013). "R: A language and environment for statistical computing. ."
571 Radwan, J., W. Babik, J. Kaufman, T. L. Lenz and J. Winternitz (2020). "Advances in the Evolutionary
572 Understanding of MHC Polymorphism." Trends in Genetics **36**(4): 298-311.
573 Rekdal, S. L., J. A. Anmarkrud, J. T. Lifjeld and A. Johnsen (2019). "Extra-pair mating in a passerine
574 bird with highly duplicated major histocompatibility complex class II: Preference for the golden
575 mean." Molecular Ecology **28**(23): 5133-5144.
576 Reusch, T. B. H., M. A. Haberli, P. B. Aeschlimann and M. Milinski (2001). "Female sticklebacks count
577 alleles in a strategy of sexual selection explaining MHC polymorphism." Nature **414**(6861): 300-302.
578 Richardson, D. S., J. Komdeur, T. Burke and T. von Schantz (2005). "MHC-based patterns of social and
579 extra-pair mate choice in the Seychelles warbler." Proceedings of the Royal Society B-Biological
580 Sciences **272**(1564): 759-767.
581 Roberts, S. C., L. M. Gosling, V. Carter and M. Petrie (2008). "MHC-correlated odour preferences in
582 humans and the use of oral contraceptives." Proceedings of the Royal Society B-Biological Sciences
583 **275**(1652): 2715-2722.
584 Schwensow, N., M. Eberle and S. Sommer (2008). "Compatibility counts: MHC-associated mate
585 choice in a wild promiscuous primate." Proceedings of the Royal Society B-Biological Sciences
586 **275**(1634): 555-564.
587 Sepil, I., R. Radersma, A. W. Santure, I. De Cauwer, J. Slate and B. C. Sheldon (2015). "No evidence for
588 MHC class I-based disassortative mating in a wild population of great tits." Journal of Evolutionary
589 Biology **28**(3): 642-654.
590 Setchell, J. M., M. J. E. Charpentier, K. M. Abbott, E. J. Wickings and L. A. Knapp (2010). "Opposites
591 attract: MHC-associated mate choice in a polygynous primate." Journal of Evolutionary Biology
592 **23**(1): 136-148.
593 Sherborne, A. L., M. D. Thom, S. Paterson, F. Jury, W. E. R. Ollier, P. Stockley, R. J. Beynon and J. L.
594 Hurst (2007). "The genetic basis of inbreeding avoidance in house mice." Current Biology **17**(23):
595 2061-2066.
596 Sin, Y. W., G. Annavi, C. Newman, C. Buesching, T. Burke, D. W. Macdonald and H. L. Dugdale (2015).
597 "MHC class II-assortative mate choice in European badgers (Meles meles)." Molecular Ecology
598 **24**(12): 3138-3150.
599 Stoffel, M. A., S. E. Johnston, J. G. Pilkington and J. M. Pemberton (2020). "Genetic architecture and
600 lifetime dynamics of inbreeding depression in a wild mammal." bioRxiv.
601 Szulkin, M., K. V. Stopher, J. M. Pemberton and J. M. Reid (2013). "Inbreeding avoidance, tolerance,
602 or preference in animals?" Trends in Ecology & Evolution **28**(4): 205-211.
603 Wedekind, C. and S. Furi (1997). "Body odour preferences in men and women: do they aim for
604 specific MHC combinations or simply heterozygosity?" Proceedings of the Royal Society B-Biological
605 Sciences **264**(1387): 1471-1479.
606 Wedekind, C., T. Seebeck, F. Bettens and A. J. Paepke (1995). "MHC-dependent mate preferences in
607 humans." Proc Biol Sci **260**(1359): 245-249.
608 Westerdahl, H. (2004). "No evidence of an MHC-based female mating preference in great reed
609 warblers." Molecular Ecology **13**(8): 2465-2470.
610 Winternitz, J. C., S. G. Minchey, L. Z. Garamszegi, S. Huang, P. R. Stephens and S. Altizer (2013).
611 "Sexual selection explains more functional variation in the mammalian major histocompatibility
612 complex than parasitism." Proceedings of the Royal Society B-Biological Sciences **280**(1769).

613 Winternitz, J. C., M. Promerova, R. Polakova, M. Vinkler, J. Schnitzer, P. Munclinger, W. Babik, J.
614 Radwan, J. Bryja and T. Albrecht (2015). "Effects of heterozygosity and MHC diversity on patterns of
615 extra-pair paternity in the socially monogamous scarlet rosefinch." Behavioral Ecology and
616 Sociobiology **69**(3): 459-469.
617 Yamazaki, K., G. K. Beauchamp, M. Curran, J. Bard and E. A. Boyse (2000). "Parent-progeny
618 recognition as a function of MHC odortype identity." Proceedings of the National Academy of
619 Sciences of the United States of America **97**(19): 10500-10502.
620 Yeates, S. E., S. Einum, I. A. Fleming, H. J. Megens, R. J. M. Stet, K. Hindar, W. V. Holt, K. J. W. Van
621 Look and M. J. G. Gage (2009). "Atlantic salmon eggs favour sperm in competition that have similar
622 major histocompatibility alleles." Proceedings of the Royal Society B-Biological Sciences **276**(1656):
623 559-566.
624 Yu, L. J., Y. G. Nie, L. Yan, Y. B. Hu and F. W. Wei (2018). "No evidence for MHC-based mate choice in
625 wild giant pandas." Ecology and Evolution **8**(17): 8642-8651.

626