First assessment of MHC diversity in wild Scottish red deer populations

Sílvia Pérez-Espona, William Paul Goodall-Copestake, Anna Savirina, Jekaterina Bobovikova, Carles Molina-Rubio, F. Javier Pérez-Barbería

S. Pérez-Espona
Royal (Dick) School of Veterinary Studies
The University of Edinburgh
Easter Bush Campus
Midlothian EH25 9RG, United Kingdom

Anglia Ruskin University
East Road, Cambridge CB1 1PT
United Kingdom

W. P. Goodall-Copestake
British Antarctic Survey
High Cross, Madingley Road
Cambridge CB3 0ET
United Kingdom

A. Savirina, J Bobovikova
Anglia Ruskin University
East Road, Cambridge CB1 1PT
United Kingdom

C. Molina-Rubio
Estación Biológica de Doñana – CSIC
Américo Vespucio s/n, Isla de la Cartuja,
E41092 Seville, Spain

F. J. Pérez-Barbería
Instituto de Investigación en Recursos Cinegéticos of CSIC-UCLM-JCCM,
Universidad de Castilla-La Mancha
Campus Universitario s/n, 02071, Albacete, Spain

Corresponding author:

Sílvia Pérez-Espona
E-mail: silvia.perez-espona@ed.ac.uk
Tel: ++44(0)131 651 7411
ORCID: 0000-0001-7098-4904

Abstract

Control and mitigation of disease in wild ungulate populations is one of the major challenges in wildlife management. Despite the importance of the Major Histocompatibility Complex (MHC) genes for immune response, assessment of diversity on these genes is still rare for European deer populations. Here, we conducted the first assessment of variation at the second exon of the MHC DRB in wild populations of Scottish highland red deer, the largest continuous population of red deer in Europe. Allelic diversity at these loci was high, with 25 alleles identified. Selection analyses indicated c. 22% of amino acids encoded found under episodic positive selection. Patterns of MHC allelic distribution were not congruent with neutral population genetic structure (estimated with 16 nuclear microsatellite markers) in the study area; the latter showing a marked differentiation between populations located at either side of the Great Glen. This study represents a first step towards building an immunogenetic map of red deer populations across Scotland to aid future management strategies for this ecologically and economically important species.

Keywords: *Cervus elaphus*, immunogenetics, Major Histocompatibility Complex, population structure, red deer, wildlife management.

78 **Introduction**

79 The impact of anthropogenic activities coupled with rapid and unprecedented climate change poses great
80 challenges for wildlife management (Altizer et al. 2003; Mawdsley et al. 2009). Among these challenges will be
81 the control of the spread of disease resulting from the emergence and re-emergence of diseases due to global
82 increases in temperature, translocations of wild and domesticated animals, and the increased contact between
83 wildlife and livestock (Daszak 2000, Mawdsley et al. 2009, Smith et al. 2006, Tompkins et al. 2015). The impacts
84 of emergence and re-emergence of disease are of particular concern for wild ungulates due to their close
85 phylogenetic relationship with different species of livestock and, hence, the higher risk of pathogens crossing
86 species barriers (Jolles and Ezenwa 2015, Martin et al. 2011, Richomme et al. 2006).
87
88 In Europe, the challenges surrounding wild ungulate management have been exacerbated by the substantial
89 increase in their distribution and numbers in the last decades, as a consequence of multiple environmental changes
90 and management practices (Milner et al. 2006 and reference therein, Apollonio et al. 2010). Importantly, the
91 predicted increase in drought and extreme temperatures is expected to affect the body condition and physiological
92 stress of individuals in some populations (East et al. 2011, Duncan et al. 2012); therefore, potentially reducing
93 immunocompetence and increasing the risk of susceptibility to infection (Patz and Reisen 2001, Acevedo-
94 Whitehouse and Duffus 2009).
95
96 It is widely recognised that genetic diversity is one of the main factors that enable populations to respond and
97 adapt to environmental change. Genetic diversity, thus, is a crucial factor to take into account for the development
98 of effective conservation and management programs for wild species (Allendorf and Luikart 2007). Among the
99 most widely used genetic loci for studying adaptation in jawed vertebrates is the Major Histocompatibility
100 Complex (MHC), a family of highly variable genes that play a central role in the immune defence response against
101 pathogen infections (Edwards and Hedrick 1998, Knapp 2005, Piertney and Oliver 2006). Polymorphism at MHC
102 loci has been shown to play an important role in the capacity of populations to fight pathogens (Lenz et al. 2009,
103 Oliver et al. 2009, Kloch et al. 2010), as well as in sexual selection and mate choice (Winternitz et al. 2013, Sin et
104 al. 2015, Santos et al. 2017), and survival fitness (Paterson et al. 1998, Pitcher and Neff 2006, Brouwer et al. 2010,
105 Eizaguirre et al. 2012). MHC molecules are group into different classes. In humans, MHC class I molecules bind
106 pathogen-derived peptides found in the cytoplasm of nucleated cells while MCH class II molecules have been
107 shown to bind pathogen-derived peptides taken from the extracellular space and encapsulated in intracellular

vesicles (Wieczorek et al. 2017). Among these MHC Class II regions, the DRB loci have been found to be the most polymorphic and, therefore, have been more widely used in studies assessing adaptive variation and resistance to pathogens in wild populations (Bernatchez and Landry 2003, Sommer 2005, Piertney and Oliver 2006, Spurgin and Richardson 2010). In particular, MHC Class II studies have focused on assessing variation at the second exon of the MHC-DRB loci, as the high allelic variation observed in this exon has been related to the ability of a population to present a wider repertoire of antigens and, subsequently, a higher potential to recognise a broader range of (extracellular) pathogens (Hughes and Hughes 1995, Reche and Reinherz 2003). Genes located between MHC Class I and Class II encoding for complement proteins in the immune response such as cytokines and heat shock proteins are referred as Class III. Furthermore, a group of genes involved in responses to inflammatory stimuli have been referred as Class IV. However, Class III and IV are not involved in the coding of antigen presentation proteins (Gruen and Weissman 2001).

Despite the importance of MHC variability for the evolutionary potential of populations to fight pathogens, and the increasing risks of re-/emergence of disease, research on the most variable MHC loci (DRB exon 2) on wild deer populations is still scarce. This gap in research is particularly notable for wild deer populations in Europe, with research currently limited to very few studies focused on single populations (Fernández de Mera et al. 2009a, Fernández de Mera et al. 2009b, Vanpé et al. 2016), populations separated by large distances (Mikko & Andersson 1995; Quéméré et al. 2015, Buczek et al. 2016) or studies that did not provide information about the specific geographical location of where samples were collected (Mikko et al. 1999). Although these studies have undoubtedly provided important insights about the diversity of MHC DRB exon 2 loci in these populations, as well as its role in parasite resistance and antler development, the geographical scale at which they were conducted precluded assessments of the effects of gene flow, a demographic process important to take into account when devising effective management strategies.

If we are to thoroughly assess the distribution of MHC allelic variation across populations to estimate potential genetic resilience to pathogens in the context of wildlife management, it is important that studies are conducted at geographical scales that allow us to estimate the influence of gene flow on the distribution of alleles across the landscape (Landry and Bernatchez 2001, Muirhead 2001, Bernatchez and Landry 2003). Here, we present the first assessment of MHC allelic variation of Scottish highland red deer (*Cervus elaphus*), one of the largest populations of red deer in Europe. This study was conducted at a geographical scale where gene flow between populations of

Scottish highland red deer could be assessed and represents the first platform to develop genetic-based protocols to inform future management strategies and monitoring of these populations. The objectives of this study were to: (i) conduct the first assessment of MHC variation in Scottish highland red deer and develop a protocol to facilitate future MHC DRB exon 2 genotyping of further populations; (ii) to conduct this study at a geographical scale that allowed us to evaluate the effects of selection and gene flow on adaptive variation by comparing patterns of allelic variation of MHC DRB exon 2 with those obtained from 16 nuclear microsatellite markers; (iii) to detect any signatures of recombination and selection on the MHC alleles found; and (iv) discuss the implications of the above on the management of Scottish highland red deer.

Methods

Study area and sampling

Male and female adult red deer (four years or older) were collected during the legal hunting seasons of 2006-2007 and 2007-2008 in four estates in the Scottish Highlands (Fig. 1), two of which were located west of the Great Glen (Tarlogie, Strathconon) and two east of the Great Glen (Inshriach, Abernethy). Major genetic differentiation was previously found on each side of this geographical feature by analysing another set of Scottish red deer populations using microsatellite markers (Pérez-Espona et al. 2008, Pérez-Espona et al. 2013) and mitochondrial DNA sequences (Pérez-Espona et al. 2009b). The estates of Strathconon and Abernethy maintain open hill red deer populations, while Tarlogie and Inshriach estates are characterised mainly by forest deer populations, *sensu* Mitchell et al. (1977). Abbreviation for estates names are as follow: STRA (Strathconon), ABNE (Abernethy), TAR (Tarlogie), INSH (Inshriach). Similar to other studies of game species or large mammals, sampling was opportunistic. In our study, samples were from individuals shot at different locations within the estate during the hunting season for each of the sexes. Average pairwise relatedness using 16 microsatellites markers (see below) and calculated in SPaGedi (Hardy and Vekemans 2002) using the pairwise relationship coefficient 'r' (Wang, 2002) estimator were as follow: STRA ($r = 0.0776 \pm 0.144$), ABNE ($r = 0.0859 \pm 0.140$), TAR ($r = -0.0009 \pm 0.136$), INSH ($r = 0.0427 \pm 0.1284$) indicating that, on average, individuals collected in a particular estate were not closely related.

MHC DRB Exon 2 genotyping

Genomic DNA for a total of 48 individuals (six males and six females per estate) was extracted from kidney tissue (stored in 100% ethanol) using the QIAGEN DNeasy Tissue Kit™, following the manufacturer's instructions. The

DNA extractions were visualized together with a DNA size marker (Hyperladder I; Bioline, UK) on a 1% agarose gel stained with GelRed™ (Biotium, US). The concentration of genomic DNA in each sample was measured with Qubit fluorometric quantitation and NanoDrop spectrophotometry (Thermo Fisher Scientific, UK) and the individual DNA extractions were subsequently diluted to the same concentration. Similar to other previous studies on Cervidae (e.g. Ditchkoff et al. 2005; Fernández de Mera et al. 2009 a,b; Kennedy et al. 2011, Cai et al. 2015, Xia et al. 2016), DNA from each individual was genotyped by amplifying exon 2 of the MHC DRB using the cattle-specific primers LA31 and LA32 (Sigurdardóttir et al., 1991). As this was the first assessment of MHC diversity in wild populations of Scottish red deer and we did not have any knowledge on the number and identity of alleles expected, two approaches were used for the genotyping: traditional cloning followed by Sanger sequencing and Roche 454 second-generation sequencing. Using these two independent methods allowed us to confirm which sequence variants represented true alleles, which is particularly important for rare alleles amplified at lower frequencies.

For Sanger sequencing, Polymerase Chain Reaction (PCR) DNA amplification was conducted in a total volume of 50µL with c.15ng of DNA template, 25µL of MyTaq™ Mix (Bioline, UK), 0.2µM of each primer, and double processed tissue culture distilled water (Sigma-Aldrich, Buchs, Switzerland) to bring the volume up to 50µL. The PCR cycling protocol involved an initial denaturation step of 94°C for 1 min, a three-step cycling of denaturing at 95°C for 15 s, annealing at 68°C for 1 min, and ramping at 0.3°C/s to an extension step of 72°C for 1 min. The cycle was repeated 33 times and was followed by a final extension of 72°C for 10 min. PCR products were visualized on 1.5% agarose gels as described above, and successful amplifications were purified using the QIAquick® PCR Purification Kit following the manufacturer's instructions (QIAGEN, UK). Purified PCR products were verified on 1.5% agarose gels and then cloned using the TOPO® TA Cloning Kit (Thermo Fisher Scientific, UK) following the manufacturer's instructions, with 6µL of iPTG/x-Gal added to each transformation to screen for recombinant plasmids. At least 56 recombinant clones per individual were picked into 10µL double distilled water. Aliquots (1.5µL) from each of these colony picks were used for PCR-insert screening using the cloning kit primers M13F and M13R, the PCR protocol for this screening followed that described above but was scaled down to a final volume of 15µL and the annealing step was reduced to 55°C. After visualizing the PCR products on 1.5 % agarose gels, a total of 48 clones per individual - that included an insert of the expected size - were sent to Source Bioscience (UK) for Sanger sequencing.

For the 454 sequencing, PCR reactions were conducted as described above for the original MHC DRB amplification, with the exception that modified versions of the primers LA31 and LA32 were used. These primers contained 5' extensions which included the 454 Lib-A adapter, key, and one of the 454 standard MID index sequences (Roche, USA). Using different combinations of MID indexes 1-7 with primer LA31 and MID indexes 1-7 with primer LA32, we amplified and uniquely MID labelled PCR products from all 48 individuals. The PCR products were purified using a QIAquick® PCR Purification Kit, visualised on 1.5% gels, and DNA concentrations were estimated as described above. All 48 PCR products were subsequently diluted to the same concentration, combined, and sent to the University of Cambridge DNA Sequencing Facility (UK) for DNA sequencing using the 454 GS Junior System (Roche, USA).

Determination of putative MHC alleles from artefacts

The software Geneious v. 9 (Biomatters, Auckland, New Zealand) was used to process the Sanger and 454 sequence data. Primer and vector sequence was trimmed from the Sanger data, then the sequences were manually edited for calling errors. For the 454 data, only high-quality reads that contained both MID index sequences were retained in order to correctly assign reads to one of the 48 individuals. Sanger and 454 sequences that contained stop codons, or which differed in length from 249 bp by more or less than a multiple of 3 bp, were discarded from further analyses as sequencing errors or potential pseudogenes. All the sequence variants per individual were checked for the presence of chimaeras using the UCHIME v 4.2.40 plug-in implemented in Geneious.

In the cloning-Sanger derived dataset, sequence variants were considered alleles only if they occurred in at least in three copies in a particular individual. Quantification of sequence variants in the 454 dataset was conducted using the software jMHC (Stuglik et al. 2011). To classify sequence variants as putative alleles or artefacts we followed a protocol similar to Herdegen et al. (2014). We calculated for each sequence variant the maximum per amplicon frequency (MPAF; Radwan et al. 2012) and then the variants were sorted according to their MPAF values. For variants with MPAF $\geq$ 1%, we checked if sequences differed by 1-2 or > 2 nucleotide sites from the more common sequence variants within an amplicon, starting with those with 1% and working upwards. Following this procedure, we found that any sequence variant with an MPAF > 4% could be considered an allele. The remaining variants, with an MPAF of 4% to 1%, were inspected on a case by case basis and those differing by > 2 nucleotide sites from the most common sequence variants were considered alleles. In order to make sure that we

did not miss any allele that might have amplified at a low frequency (< 1%), we compared all alleles found for each individual using both sequencing approaches.

MHC diversity and population structure

Alleles identified in this study were confirmed to be red deer MHC DRB exon 2 alleles using the megablast search algorithm implemented in Geneious. Sequence polymorphism and the average number of nucleotide differences between alleles (k) for each of the populations were calculated using the software DnaSP v. 5 (Librado and Rozas 2009). Further genetic diversity measures were calculated using GenAlEx v. 6.502 (Peakall and Smouse 2012) and FSTAT (Goudet 1995). GenAlEx was also used to identify private alleles. MHC population structure in the study area was assessed by performing hierarchical AMOVAs with populations nested into regions (i.e west vs east of the Great Glen) in GenAlEx. Population structure was further analysed with the Bayesian clustering-based method implemented in the software STRUCTURE v. 2.3.4 (Pritchard et al. 2000). Each allele was considered a separate dominant locus and the data was coded as binary indicating presence (1) or absence (0) of the allele in a particular individual. The most likely number of genetic populations (K) was estimated by conducting five independent runs for K = 1-10 using a burn-in of 500,000 replications, 10^6 Markov chain Monte Carlo steps and assuming a model of admixture and a model of correlated of frequencies among populations. The software STRUCTURE HARVESTER Web v. 0.6.94 (Earl and VonHoldt 2011) was used to calculate Evanno's ΔK (Evanno et al. 2005). Further visualization of STRUCTURE plots was conducted in DISTRUCT version 1.1 (Rosenberg 2004). Furthermore, we conducted a linear discriminant analysis (LDA; Venables and Ripley 2002) by finding the linear combination of the total number of MHC alleles that best characterised this allelic diversity in our four red deer population samples. LDA is a multivariate statistical tool that explicitly attempts to model the difference between classes of data in order to assess how good a particular classification in k distinct groups is given a set of predictors (in our study west/east of the Great Glen, forest/open hill habitat, sex or population). Statistical analyses and graphical representation of results were conducted in R using the packages MASS (Venables and Ripley 2002), Adehabitat (Calenge 2006), and ggplot2 (Wickham 2009).

Microsatellite genotyping

A total of 96 individuals (12 females and 12 males from each estate) from the four estates included in this study were genotyped at 16 microsatellite markers (CP26, FCB5, FCB304, JP38, RT1, RT7, TGLA94, RT25, BM757, RM188M, T156, T26, T501, T193, BM888, RT13) using the primers and procedures as described in Pérez-Espona

et al. (2008). Multiplex PCR products were sent to Source Bioscience (UK) for fragment analyses on an ABI 3730 capillary sequencer (Applied Biosystems, USA) together with the internal size standard GeneScan 500 LIZ (Applied Biosystems). Fragment analysis was conducted using the software GeneMapper™ v. 3.0 (Applied Biosystems).

Microsatellite diversity and population structure

Deviations from Hardy-Weinberg equilibrium (HWE) for each estate, tests for linkage disequilibrium (LD) across all pairs of loci, and measures of genetic diversity were conducted using the software FSTAT. Genetic diversity analyses and the presence of private alleles were estimated in GenAlEx. Population genetic structure was assessed by conducting hierarchical AMOVAs in GenAlEx using the same parameters as for the MHC allelic data. Population genetic structure was further estimated with the software STRUCTURE v. 2.3.4 (Pritchard et al. 2000) using an identical approach to that used for the MHC dataset but setting the analyses for codominant markers (see above).

Detecting signatures of recombination and selection on MHC alleles

Tests of recombination and selection were conducted using the open-source software package Hyphy (Hypothesis Testing using Phylogenies, www.hyphy.org). We first tested for evidence for recombination at multiple breakpoints using GARD (Genetic Algorithm Recombination Detection; Kosakovsky Pond *et al.* 2006) prior to any selection test, as selection tests are sensitive to recombination in the dataset. A Mixed Effects Model of Evolution (MEME; Murrell et al. 2012) was performed in order to test for pervasive and episodic positive selection on individual codon sites. Any potential positively selected site was then compared with the location of nucleotide positions coding for amino acids within the Peptide Binding Region (PBR) in the equivalent human MHC molecule by aligning the red deer MHC allele sequences to the human MHC sequence in Reche and Reinherz (2003).

Results

Genotyping of MHC alleles in Scottish red deer

From the 2,304 Sanger sequences generated (48 clones sequenced per individual), the average number of sequence variants per individual, after removing bad quality sequences and potential pseudogenes (sequences containing stop codons), was 39.31 ± 8.15 (range: 14 to 48). No chimaeras were identified in the remaining sequences after

conducting UCHIME analyses for each of the individuals. Sequencing using the 454 approach yielded 82,278 sequences assigned to individuals with an average of 1,714 ± 605.47 sequences per individual (range: 769 to 2,950). From these sequences, 15.49% ± 11.73 were discarded as their length was not multiple of 3 of the expected fragment size (249bp). A further 27.65% ± 10.07 were discarded due to sequences containing stop codons (and therefore indicating potential pseudogenes). UCHIME analyses indicated a lack of presence of chimaeras in the remaining sequences. After this sequence filtering, the average number of sequences per individual was 393 ± 106 (range: 128-619). The total number of MHC DRB exon 2 alleles found in this study was 25 (see below). All of these alleles were found with both methodologies (cloning-Sanger sequencing and 454 sequencing); however, three rare alleles would have been missed by the Sanger sequencing approach, as they were represented by only one sequence (rather than at least three).

Genetic diversity and population structure

MHC dataset

Allelic sequences have been deposited in DDBJ (DNA Data Bank of Japan) with accession numbers LC379925-LC379949, with the alleles named Ceel-DRB*1 – Ceel-DRB*25. In terms of sequence variation, 78 out of 249 sites were variable, of which 64 were parsimony informative and 14 singletons. The overall mean distance between the 25 alleles was k = 24.62. Sequence polymorphism and overall mean distance between alleles were similar between the four red deer populations studied (Table 1). Translation of the nucleotide sequences into amino acids also resulted in 25 unique sequences (Fig. S1). Nucleotide BLAST searches optimising for high similarity sequence (megablast) confirmed that all the allele sequences were characteristic of *Cervus* MHC DRB exon 2. Out of the 25 alleles found, 21 were novel (i.e. not previously found in previous studies of ungulates). The remaining four alleles (Ceel-DRB*3, Ceel-DRB*5, Ceel-DRB*16, Ceel-DRB*20) matched 100% with alleles previously found in a managed population of Iberian red deer (Accession numbers: EU573264, EU573277, EU573285, EU573271, respectively; Fernández de Mera et al. 2009b). Ceel-DRB*16 also matched 100% with an allele found in Ussuri sika (Accession number: AY679505; Wu et al. unpublished).

In our study area, the number of alleles per individual ranged from 1 to 6 (mean 2.81 ± 1.32 SD). Therefore, suggesting at least three MHC loci in the Scottish highland red deer populations. Out of the 48 Scottish red deer genotyped, 10 individuals were homozygous, with seven of them being homozygous for the most common allele in the study area (Ceel-DRB*2). No homozygous individuals were found in STRA. Ten of the alleles were shared

between the four populations at different frequencies, with 4-5 alleles relatively common across the populations, in particular, the most common allele (Ceel-DRB*2, found in 52% of the individuals). The rest of the alleles were found at much lower frequencies, with private alleles in ABNE, STRA, and TAR (Fig. 2).

Population structure estimated with hierarchical AMOVA indicated that most of the genetic variation was attributable to within-population differences (94%), although this estimate was not statistically significant (Table 2). Differences among populations within regions accounted for 6% of the genetic variation (P = 0.036), with no genetic variation explained due to differences among regions. Although the AMOVA indicated genetic differentiation between populations, the STRUCTURE results showed that this was not congruent with spatial structure (Fig. 3). The analyses supported that K = 3 was the most likely number of genetic populations in the study area (Fig. S2). However, inspection of the resulting plot indicated high levels of admixture within all the populations, with more similarity between the populations STRA and INSH.

The proportion of trace of the first three dimensions (LD1:3) of the LDA against population were LD1 = 0.5507, LD2 = 0.3403 and LD3 = 0.1090. The coefficients of the linear discriminants of the first two dimensions together with the convex hull for each population were plotted to graphically represent the degree of overlap between populations (Fig. 4). The first two dimensions explained 89% of the variance of the 25 alleles in relation to the four populations. The discriminant scores of the first dimension characterised two clear groups of populations [INSH, STRA] and [TAR, ABNE]; however, this first dimension was less efficient at separating INSH from STRA and had little power to discriminate between TAR and ABNE (Fig. 4). The discriminant scores of the second dimension characterised two groups of populations: ABNE and [TAR, INSH, STRA]. Furthermore, we took a regression approach using the scores of each of the linear discriminants against the four populations as levels of the treatment in the analysis. The coefficients of the regressions were used to calculate pair-wise contrasts between populations to assess if they could be differentiated. The contrasts for each linear discriminant are displayed in Table S1 (Supplementary Material). The results corroborate the visual separation of populations of Figure 4.

Microsatellite dataset

No departures from Hardy-Weinberg equilibrium and no linkage disequilibrium were detected in any of the sampling sites or pairs of loci analysed. The number of alleles and estimates of H_O and H_E were similar between the four estates; however, allelic richness was slightly higher in TAR and slightly lower in STRA (Table 1). Private

alleles were found in all four estates but with a higher frequency in TAR and ABNE (Fig. 5). Population differentiation estimated with hierarchical AMOVA (Table 2) indicated that although most of the genetic variation was found within populations (92%; P = 0.001), some of the variation was attributable to regions (4%, P = 0.001) and among populations within regions (4%; P = 0.001). Analyses in STRUCTURE and Evanno's Delta indicated that the most likely number of genetic clusters was K = 2 (Fig. S3), corresponding with the major genetic differentiation between estates located west and east of the Great Glen (Fig. 3), although with gene flow between these two main geographical areas. The gene flow was found to be predominantly west to east, with individuals sampled from INSH being more genetically admixed that those from ABNE.

Signatures of recombination and natural selection in MHC DRB in Scottish highland red deer

The screening of alignments conducted by GARD did not detect any evidence for recombination breakpoints. Positive selection was identified for codon positions 1, 3, 5, 20, 36, 49, 52, 66, 70, 78 (Fig. S1). Eight of these codons coincided with PBR sites in humans, the other two were located in within the proximity of other human PBR sites (Fig. S1). Three amino acid positions identified as PBR in humans (codons 74, 77, and 81) were not variable in our Scottish highland red deer dataset. Six of the remaining positions identified as PBR in humans, although not found to be under positive selection, were variable (two or three amino acids variants) in our red deer dataset.

Discussion

MHC variation in Scottish highland red deer

This first assessment of MHC diversity in Scottish red deer populations revealed a total of 25 MHC alleles. MHC studies on deer are not directly comparable due to differences in methodological approaches (experimental design and approaches to genotyping) but also due to intrinsic differences between the studied populations (population size, population history, demography, management). Nonetheless, if we consider the number of individuals analysed and the geographical scale at which our study was conducted, MHC DRB exon 2 variation in Scottish highland red deer was found to be larger than in previous studies of wild populations of Cervidae populations from Europe (Mikko and Andersson 1995; Mikko et al. 1999; Fernández de Mera et al. 2009a, Fernández de Mera et al. 2009b, Buczek et al. 2016, Quéméré et al. 2015, Vanpé et al. 2016), North America (Ditchkoff et al. 2001, Ditchkoff et al. 2005, Kennedy et al. 2011, Van Den Bussche et al. 2002) and Asia (Cai et al. 2015, Yao et al.

2015; Xia et al. 2015; Table 3). Indeed, out of the 25 alleles found in this study, 21 alleles have not been reported in any previous MHC DRB exon 2 assessments in other species or populations.

Four of the MHC DRB exon 2 alleles found in this study were also found in Spanish red deer or Ussari sika deer. This sharing can be explained by trans-species polymorphism – the persistence of allelic lineages from common ancestors to descendant species (Klein et al., 1998). However, we should not discard the effect of past deer management practices as translocations of Scottish red deer into other European countries. Introductions of exotic deer in Scotland are well documented (Whitehead 1960, Whitehead 1964, Pérez-Espona et al. 2009a) and could be a potential factor increasing MHC variability in Scottish red deer. Previous genetic studies have confirmed a low effect of these introductions on the genetic makeup of Scottish highland red deer (Pérez-Espona et al. 2009b, Pérez-Espona et al. 2011, Pérez-Espona et al. 2013, Smith et al. 2018), with the exception of extensive red deer and sika deer hybridisation detected in South Kyntire (Senn and Pemberton 2009, Senn et al. 2010a, Senn et al. 2010b, Smith et al. 2018) and some evidence of hybridisation between these two deer species in the North Highlands (three out of 568 individuals surveyed confirmed as hybrids; Smith et al., 2018). The MHC DRB exon 2 allele shared with sika deer and Spanish red deer (allele Ceel-DRB*16) was found in three males from Strathconon and one female from Abernethy. In Strathconon, sika deer presence has been regularly reported in the past years and sika are known to be established in nearby forests (Seivwright 2017), with some evidence of hybridisation between red and sika deer in the nearby estate in Torrachilty (Smith et al. 2018). The presence of sika deer in Abernethy is rare but they are expanding their range in this area; however, no evidence of hybridisation in this or nearby estates was found in the study by Smith et al. (2018). Incoming alleles previously not present in a population could be selected for if they confer a selective advantage, and introgress more rapidly than neutral alleles (Schierup et al. 2000, Barton 2001, Muirhead 2001) in particular in large populations (Kimura and Ohta 1969). Therefore, despite the absent or very low levels of hybridisation found in nearby areas of Strathconon and Abernethy, further studies should investigate more thoroughly the potential effect of hybridisation on MHC allelic diversity in these populations.

Comparisons of the MHC alleles found in our study to those found in a previous study on farmed red deer in New Zealand (Swarbrick et al. 1995), potentially including individuals descendant from Scottish red deer populations, were difficult due to the lack of complete overlap between sequences (i.e. Swarbrick et al.'s sequences start 24 bp downstream than any other available sequences for MHC DRB exon 2 sequences in ungulates). However, after

406 trimming the sequences to an overlapping fragment length of 225bp, three of the alleles found in our study (Ceel-
407 DRB*6, Ceel-DRB*14, Ceel-DRB*20) were found to match 100% with three alleles found in the New Zealand
408 farmed red deer.

409 **Genetic diversity and population structure**

410 Measures of genetic differentiation estimated with neutral loci are important to infer demographic processes
411 affecting populations, such as dispersal and population history, and, thus, are important to define conservation or
412 management units (Palsbøll et al. 2007). In our study, genetic diversity values obtained with microsatellite data
413 were high (H_E = 0.755-0.812; Allelic richness = 6.83-7.69) and similar to those found in previous studies of
414 mainland Scottish red deer (Pérez-Espona et al. 2008; Pérez-Espona et al. 2010; Pérez-Espona et al. 2013). Genetic
415 differentiation using the microsatellite dataset was concordant with the geographical location of the populations
416 and the effect of landscape features on the Great Glen; previously shown to be a barrier to Scottish mainland red
417 deer gene flow (Pérez-Espona et al., 2008; Pérez-Espona et al., 2009b; Pérez-Espona et al., 2013). Neutral loci,
418 however, do not provide information about the patterns of adaptive variation across the landscape which is crucial
419 to devise management strategies in the context of emergence or re-emergence and spread of disease (Hedrick et
420 al. 2001, Funk et al. 2012). The population structure analyses of the MHC dataset indicated that although structure
421 was found between populations, the patterns of differentiation were not concordant with geography; with no
422 differentiation found between populations located at either side of the Great Glen. Furthermore, the STRUCTURE
423 results showed high levels of admixture within the populations. These results indicate that patterns of MHC
424 variation in the study area are not mainly due to gene flow between populations and, therefore, that balancing
425 selection might have an effect on the distribution of MHC allelic variation among populations (Hedrick 1999,
426 Schierup et al. 2000). Patterns of MHC polymorphism were not explained by differences in habitat (open hill
427 versus forested), indicating that similar pathogen-driven selection pressures might be acting on the studied
428 populations. The action of balancing selection was further supported by the analyses of selection on MHC
429 diversity over evolutionary time, with approximately 22% of the amino acids of the MHCDRB exon 2 in Scottish
430 highland red deer were found to be under episodic positive selection. Eight of the codon positions identified under
431 positive selection coincided with 19 of the PBR sites described by X-ray crystallography in humans (Reche and
432 Reinherz 2003), but our results indicated that other codons (differing from those found in humans) are likely to be
433 involved for peptide binding and subsequent immunological response in red deer.
434

435 **Implications for management**

Our study provided the first insights into MHC diversity in Scottish highland red deer, one of the largest populations of red deer in Europe. The thorough and successful approach to genotyping MHC alleles taken in our study lays the foundation for future studies of MHC diversity in red deer populations across the Scottish mainland and islands. Large congruence in the identification of MHC alleles between the traditional cloning-Sanger sequencing and 454 second-generation sequencing methods, confirmed that future studies could rely on the use of next generation sequencing for the identification of MHC allelic diversity, as these modern sequencing methods (e.g. 454, MiSeq, Ion Torrent, Nanopore) offer a more time and cost-effective protocol for genotyping MHC DRB exon 2 in red deer.

The management-relevant scale at which our study was conducted allowed us to compare the potential influence of demographic processes such as gene flow on the spatial distribution of MHC allelic variation. The main red deer management strategies in Scotland are organised in Deer Management Groups (DMGs). The delimitation of these groups is, generally, set by taking into account natural or/and man-made geographical features that might restrict deer movement across the landscape. A previous genetic study using microsatellite markers supported this management approach, with landscape features having a significant effect as barriers or facilitators to gene flow between Scottish highland red deer populations (Pérez-Espona et al. 2008). However, spatial patterns of MHC diversity in our study area were not concordant with those found with microsatellite markers; indicating that the delimitation of Deer Management Groups might not reflect the immunogenetic variation across Scotland. Therefore, other units of management will be required for devising effective strategies towards the control of the emergence or spread of disease in Scotland. In this context, it would be of great benefit that the genetic approach adopted in our study is expanded to other areas in Scotland so that an 'immunogenetic map' of red deer populations can be generated. This could be attained by genotyping individuals for MHC loci and, ideally, for other immunogenetic loci (Acevedo-Whitehouse and Cunningham 2006, Quéméré et al. 2015) and candidate genes associated with particular diseases. Further assessments of immunogenetic variability in a larger number of red deer populations would, therefore, facilitate rigorous tests on the association of immunogenetic loci and body condition data for red deer individuals. This type of information, together with data on the spatial distribution of neutral genetic variation, would be crucial for an effective and long-term sustainable management of Scottish red deer populations (McKnight et al. 2017).

Acknowledgements

Deer stalkers and deer managers of the estates of Tarlogie, Strathconon, Inshriach and Abernethy are greatly thanked for the collection of samples. A. Jones, K. Russell, S. Joinson, and J. Hennessy are thanked for assistance with microsatellite genotyping and S. Requena (CSIC) for map reproduction. Cambridge Conservation Forum and the Cambridge Conservation Initiative are thanked for allowing Sílvia Pérez-Espona to use their office space at the David Attenborough Building while preparing this manuscript. This study was funded by the British Deer Society and samples were obtained from a project funded through Rural Affairs Food and Environment Strategic Research - Scottish Government.

References

Acevedo-Whitehouse, K., Cunningham, A.A., 2006. Is MHC enough for understanding wildlife immunogenetics? Trends Ecol. Evol. 21, 433–438.

Acevedo-Whitehouse, K., Duffus, A.L.J., 2009. Effects of environmental change on wildlife health. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364, 3429–3438.

Allendorf, F.W., Luikart, G., 2007. Conservation and the Genetics of Populations, Management. Wiley-Blackwell.

Altizer, S., Harvell, D., Friedle, E., 2003. Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol. Evol. 18, 589–596.

Apollonio, M., Andersen, R., Putman, R., 2010. European Ungulates and their Management in the 21st Century. Cambridge University Press, Cambridge.

Barton, N.H., 2001. The role of hybridization in evolution. Mol. Ecol. 10, 551–568.

Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? J. Evol. Biol. 16, 363–377.

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

Buczek, M., Okarma, H., Demiaszkiewicz, A.W., Radwan, J., 2016. MHC, parasites and antler development in red deer: no support for the Hamilton & Zuk hypothesis. J. Evol. Biol. 29, 617–632.

Cai, R., Shafer, A.B.A., Laguardia, A., Lin, Z., Liu, S., Hu, D., 2015. Recombination and selection in the major histocompatibility complex of the endangered forest musk deer (*Moschus berezovskii*). Sci. Rep. 5, 17285.

Calenge, C., 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. Ecol. Modell. 197, 516–519.

Daszak, P., 2000. Emerging infectious diseases of wildlife - threats to biodiversity and human health. Science. 287, 443–449.

Ditchkoff, S.S., Hoofer, S.R., Lochmiller, R.L., Masters, R.E., Van Den Bussche, R. a., 2005. MHC-DRB evolution provides insight into parasite resistance in white-tailed deer. Southwest. Nat. 50, 57–64.

Ditchkoff, S.S., Lochmiller, R.L., Masters, R.E., Hoofer, S.R., Van Den Bussche, R. a, 2001. Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): evidence for good-genes advertisement. Evolution. 55, 616–25.

Duncan, C., Chauvenet, A.L.M., McRae, L.M., Pettorelli, N., 2012. Predicting the future impact of droughts on ungulate populations in arid and semi-arid environments. PLoS One 7, e51490.

Earl, D.A., VonHoldt, B.M., 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. 4, 359–361.

East, M.L., Bassano, B., Ytrehus, B., 2011. The role of pathogens in the population dynamics of European ungulates, in: Ungulate Management in Europe: Problems and Practices. Cambridge University Press, Cambridge, pp. 319–348.

Edwards, S. V., Hedrick, P.W., 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. 13, 305–311.

Eizaguirre, C., Lens, T.L., Kalbe, M., Milinski, M., 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat. Comm. 3, 261.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.

Fernández de Mera, I.G., Vicente, J., Naranjo, V., Fierro, Y., Garde, J.J., de la Fuente, J., Gortázar, C., 2009. Impact of major histocompatibility complex class II polymorphisms on Iberian red deer parasitism and life history traits. Infect. Genet. Evol. 9, 1232–1239.

Fernández de Mera, I.G., Vicente, J., Pérez de la Lastra, J.M., Mangold, A.J., Naranjo, V., Fierro, Y., de la Fuente, J., Gortázar, C., 2009. Reduced major histocompatibility complex class II polymorphism in a hunter-managed isolated Iberian red deer population. J. Zool. 277, 157–170.

Funk, W.C., McKay, J.K., Hohenlohe, P.A., Allendorf, F.W., 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27, 489–496.

Goudet, J., 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. J. Hered. 86, 485–486.

Gruen, J., Weissman, S.M. 2001. Human MHC class III and IV genes and disease associations. Front. Biosci. 1, 6:D960-972.

Hardy, O.J., Vekemans, X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2, 618-620.

Hedrick, P., Parker, K., Lee, R., 2001. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. Mol. Ecol. 10, 1399–1412.

Hedrick, P.W., 1999. Perspective : highly variable loci and their interpretation in evolution and conservation. Evolution. 53, 313–318.

Herdegen, M., Babik, W., Radwan, J., 2014. Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. J. Evol. Biol. 27, 2347–2359.

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

Jolles, A.E., Ezenwa, V.O., 2015. Ungulates as model systems for the study of disease processes in natural populations. J. Mammal. 96, 4–15.

Kennedy, L.J., Modrell, A., Groves, P., Wei, Z., Single, R.M., Happ, G.M., 2011. Genetic diversity of the major histocompatibility complex class II in Alaskan caribou herds. Int. J. Immunogenet. 38, 109–119.

Kimura, M., Ohta, T., 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61, 763–771.

Klein, J., Sato, A., Nagl, S., O'hUigín, 1998. Molecular trans-species polymorphism. Annu. Rev. Ecol. Syst. 29, 1–21.

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

Knapp, L.A., 2005. The ABCs of MHC. Evol. Anthropol. 37, 28–37.

Kosakovsky Pond, S.L., Posada, D., Gravenor, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23, 1891–1901.

Landry, C., Bernatchez, L., 2001. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Mol. Ecol. 10, 2525–2539.

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

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Martin, C., Pastoret, P.-P., Brochier, B., Humblet, M.-F., Saegerman, C., 2011. A survey of the transmission of infectious diseases/infections between wild and domestic ungulates in Europe. Vet. Res. 42, 70.

Mawdsley, J.R., Malley, R.O., Ojima, D.S., 2009. A review of climate-change adaptation strategies for wildlife management and biodiversity conservation. Conserv. Biol. 23, 1080–1089.

McKnight, D.T., Schwarzkopf, L., Alford, R.A., Bower, D.S., Zenger, K.R., 2017. Effects of emerging infectious diseases on host population genetics: a review. Conserv. Genet. 18, 1235–1245.

Mikko, S., Andersson, L., 1995. Low major histocompatibility complex class II diversity in European and North American moose. Proc. Natl. Acad. Sci. USA 92: 4259-4263.

Mikko, S., Røed, K., Schmutz, S., Andersson, L., 1999. Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. Immunological Reviews 167: 169-178.

Milner, J.M., Bonenfant, C., Mysterud, A., Gaillard, J.-M., Csányi, S., Stenseth, N.C., 2006. Temporal and spatial development of red deer harvesting in Europe: biological and cultural factors. J. Appl. Ecol. 43, 721–734.

Mitchell, B., Staines, B., Welch, D., 1977. Ecology of red deer. A research review relevant to their management in Scotland. Cambridge, United Kingdom.

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

577 Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S.L., 2012. Detecting
578 individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.
579 Oliver, M.K., Telfer, S., Piertney, S.B., 2009. Major histocompatibility complex (MHC) heterozygote
580 superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). Proc. R. Soc. B Biol.
581 Sci. 22, 1119–1128.
582 Palsbøll, P.J., Berube, M., Allendorf, F.W., 2007. Identification of management units using population genetic
583 data. Trends Ecol. Evol. 22, 11–16.
584 Paterson, S., Wilson, K., Pemberton, J.M., 1998. Major histocompatibility complex variation associated with
585 juvenile survival and parasite resistance in a large unmanaged ungulate population. Proc. Natl. Acad. Sci.
586 U. S. A. 95, 3714–3719.
587 Patz, J.A., Reisen, W.K., 2001. Immunology, climate change and vector-borne diseases. Trends Immunol. 22,
588 171–172.
589 Peakall, R.., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for
590 teaching and research - an update. Bioinformatics 28, 2537–2539.
591 Pérez-Espona, S., Hall, R.J., Pérez-Barbería, F.J., Glass, B.C., Ward, J.F., Pemberton, J.M., 2013. The impact of
592 past introductions on an iconic and economically important species, the red deer of Scotland. J. Hered.
593 104, 14–22.
594 Pérez-Espona, S., Pemberton, J.M., Putman, R., 2009a. Red and sika deer in the British Isles, current
595 management issues and management policy. Mamm. Biol. 74, 247–262.
596 Pérez-Espona, S., Pérez-Barbería, F.J., Goodall-Copestake, W.P., Jiggins, C.D., Gordon, I.J., Pemberton, J.M.,
597 2009b. Genetic diversity and population structure of Scottish Highland red deer (*Cervus elaphus*)
598 populations: a mitochondrial survey. Heredity. 102, 199–210.
599 Pérez-Espona, S., Pérez-Barbería, F.J., Goodall-Copestake, W.P., Jiggins, C.D., Gordon, I.J., Pemberton, J.M.,
600 2010. Variable extent of sex-biased dispersal in a strongly polygynous mammal. Mol. Ecol. 19, 3101-
601 3113.
602 Pérez-Espona, S., Pérez-Barbería, F.J., Mcleod, J.E., Jiggins, C.D., Gordon, I.J., Pemberton, J.M., 2008.
603 Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). Mol. Ecol. 17, 981–
604 996.
605 Pérez-Espona, S., Pérez-Barbería, F.J., Pemberton, J.M., 2011. Assessing the impact of past wapiti introductions
606 into Scottish Highland red deer populations using a Y chromosome marker. Mamm. Biol. 76, 640–643.
607 Piertney, S.B., Oliver, M.K., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity.
608 96, 7–21.
609 Pitcher, T.E., Neff, B.D., 2006. MHC class IIB alleles contribute to both additive and nonadditive genetic effects
610 on survival in Chinook salmon. Mol. Ecol. 15, 2357–2365.
611 Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data.
612 Genetics 155, 945–959.
613 Quéméré, E., Galan, M., Cosson, J.F., Klein, F., Aulagnier, S., Gilot-Fromont, E., Merlet, J., Bonhomme, M.,
614 Hewison, A.J.M., Charbonnel, N., 2015. Immunogenetic heterogeneity in a widespread ungulate: the
615 European roe deer (*Capreolus capreolus*). Mol. Ecol. 24, 3873–3887.
616 Radwan, J., Zagalska-Neubauer, M., Cichon, M., Sendecka, J., Kulma, K., Gustafsson, L., Babik, W., 2012.
617 MHC diversity, malaria and lifetime reproductive success in collared flycatchers. Mol. Ecol. 21, 2469–
618 2479.
619 Reche, P.A., Reinherz, E.L., 2003. Sequence variability analysis of human class I and class II MHC molecules:
620 functional and structural correlates of amino acid polymorphisms. J. Mol. Biol. 331, 623–641.
621 Richomme, C., Gauthier, D., Fromont, E., 2006. Contact rates and exposure to inter-species disease transmission
622 in mountain ungulates. Epidemiol. Infect. 134, 21–30.
623 Rosenberg, N.A., 2004. Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes 4,
624 137–138.
625 Santos, P.S.C., Michler, F.W., Sommer, S., 2017. Can MHC-assortative partner choice promote offspring
626 diversity? A new combination of MHC-dependent behaviours among sexes in a highly successful invasive
627 mammal. Mol. Ecol. 26, 2392–2404.
628 Schierup, M.H., Vekemans, X., Charlesworth, D., 2000. The effect of subdivision on variation at multi-allelic
629 loci under balancing selection. Genet. Res. (Camb). 76, 51–62.
630 Seivwright, L., 2017. Strathconon Deer Management Group Part 1: Deer Management Plan Information &
631 Public Interest Actions.
632 Senn, H.V., Goodman, S.J., Swanson, G.M., Abernethy, K.A., Pemberton, J.M., 2010a. Investigating temporal
633 changes in hybridization and introgression in a predominantly bimodal hybridizing population of invasive
634 sika (*Cervus nippon*) and native red deer (*C. elaphus*) on the Kintyre Peninsula. Mol. Ecol. 19, 910–924.
635 Senn, H.V., Pemberton, J.M., 2009. Variable extent of hybridization between invasive sika (*Cervus nippon*) and
636 native red deer (*C. elaphus*) in a small geographical area. Mol. Ecol. 18, 862–876.

Senn, H.V., Swanson, G.M., Goodman, S.J., Barton, N.H., Pemberton, J.M., 2010b. Phenotypic correlates of hybridisation between red and sika deer (genus *Cervus*). J. Anim. Ecol. 79, 414–425.
Sigurdardóttir, S., Borsch, C., Gustafsson, K., Anderson, L., 1991. Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. Anim. Genet. 22, 199–209.
Sin, Y.W., Annavi, G., Newman, C., Buesching, C., Burke, T., Macdonald, D., Dugdale, H.L., 2015. MHC class II-assortative mate choice in European badgers (*Meles meles*). Mol. Ecol. 24, 3138–3150.
Smith, K.F., Sax, D.F., Lafferty, K.D., 2006. Evidence for the role of infectious disease in species extinction and endangerment. Conserv. Biol. 20, 1349–1357.
Smith, S.L., Senn, H. V., Pérez-Espona, S., Wyman, M.T., Heap, E., Pemberton, J.M., 2018. Introgression of exotic Cervus (*nippon* and *canadensis*) into red deer (*Cervus elaphus*) populations in Scotland and the English Lake District. Ecol. Evol. 8, 2122–2134.
Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2, 16.
Spurgin, L.G., Richardson, D.S., 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. R. Soc. B Biol. Sci. 277, 979–988.
Stuglik, M.T., Radwan, J., Babik, W., 2011. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Mol. Ecol. Resour. 4, 739–742.
Swarbrick, P.A., Schwaiger, F.W., Eppen, J.T., Buchan, G.S., Griffin, J.F., Crawford, A.M., 1995. Immunogenetics 42, 1-19.
Tompkins, D.M., Carver, S., Jones, M.E., Krkošek, M., Skerrat, L.F., 2015. Emerging infectious diseases of wildlife: a critical perspective. Trends Parasitol. 31, 149–159.
Van Den Bussche, R.A., Ross, T.G., Hoofer, S.R., 2002. Genetic variation at a major histocompatibility locus and among populations of white-tailed deer (*Odocoileus virginianus*). J. Mammal. 83, 31–39.
Vanpé, C., Debeffe, L., Galan, M., Hewison, A.J.M., Gaillard, J.-M., Gilot-Fromont, E., Morellet, N., Verheyden, H., Cosson, J.F., Cargnelutti, B., Merlet, J., Quéméré, E., 2016. Immune gene variability influences roe deer natal dispersal. Oikos 125, 1790–1801.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with R, Fourth. ed. Springer.
Wang, J., 2002. An estimator for pairwise relatedness using molecular markers. Genetics.
Whitehead, G.K., 1964. The deer of Great Britain and Ireland. Routledge & Kegan Paul, London.
Whitehead, G.K., 1960. The deer stalking grounds of Great Britain and Ireland. Hollis and Carter, London.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, New York.
Wieczorek, M., Abualrous, E.T., Sticht, J., Álvaro-benito, M., Werner, J.M., 2017. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins : Conformational Plasticity in Antigen Presentation. Front. Immunol. 8, 1–16.
Winternitz, J.C., Minchey, S.G., Garamszegi, L.Z., Huang, S., Stephens, P.R., Altizer, S., 2013. Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. Proc. R. Soc. B Biol. Sci. 280, 20131605.
Xia, S., Fan, Z., Zhang, X., Jie, C., Zhang, X., Yue, B., 2016. Molecular polymorphism of MHC-DRB gene and genetic diversity analyses of captive forest musk deer (*Moschus berezovskii*). Biochem. Syst. Ecol. 67, 37-43.
Yamazaki, K., Beauchamp, G.K., 2007. Genetic basis for MHC-dependent mate choice. Adv. Genet. 59, 129–145.
Yao, G., Zhu, Y., Wan, Q.-H., Fang, S.-G., 2015. Major histocompatibility complex class II genetic variation in forest musk deer (*Moschus berezovskii*). Anim. Genet. 46, 535–543.

Figure legends

Figure 1. Map indicating the location of the estates sampled at either side of the Great Glen. The discontinuous line indicates delimitation of the Cairngorms National Park.

Figure 2. Plot of the frequency of MHC DRB exon 2 alleles in the four populations of Scottish highland red deer. Ten alleles were found in all populations at different frequencies. The most common allele was CeelDRB*2 which was present in c. 52% of the individuals. Ten alleles were private, only found in a particular population.

Figure 3. Plot of genetic structure inferred using the MHC (top) and microsatellite (bottom) data sets in terms of estimates of Q (estimated membership coefficient for each individual) for the selected K. Vertical lines are broken into coloured segments showing the proportion of each individual's genotype assigned to each of the inferred K.

Figure 4. Plot of MHC DRB exon 2 variance expressed as the coefficients of the linear discriminant of the first two dimensions and convex hull for each of the Scottish highland red deer populations. The first two dimensions explained 89% of the variance of the 25 alleles in relation to the four populations. Symbols indicate data from populations; dots: ABNE, crosses: TAR, triangles: INSH, squares: STRA.

Figure 5. Plot of the average number of alleles and private alleles for the 16 microsatellite markers used to genotype Scottish highland red deer. The average number of alleles were similar between the populations. Private alleles were slightly higher in TAR and ABNE.

SUPPORTING INFORMATION

Figure S1. Amino acid composition of the 25 MHC DRB exon 2 alleles found in Scottish highland red deer.

Figure S2. Results from STRUCTURE for the analyses of population structure using MHC DRB exon 2 loci.

Figure S3. Results from STRUCTURE for the analyses of population structure using 16 microsatellite loci.

Table S1. Contrasts of the estimates of the regression analyses of each of the first three linear discriminants against populations. Significant p-values indicate differences between pairs of populations for the corresponding linear discriminant. The results are consistent with Figure 4.

Table S2. Summary of MHC allelic variation studies in Cervidae. Please note that studies only assessing variation of MHC DRB expressed loci are not included in this table. Superscript numbers near the species indicate the publication associated with the study.

740 **Table 1.** Genetic diversity indices within each population of Scottish highland red deer for MHC DRB exon 2 and microsatellite data
741
742

	MHC DRB exon 2 data							Microsatellite data			
Population	**Na**	**C**	**v**	**Pi**	**S**	**k**	**π**	**Mean Na/locus**	**Allelic richness**	**H_O ± SD**	**H_E ± SD**
TAR	16	178	71	55	16	24.567	0.099 ± 0.006	8.938± 2.05	7.694	0.777± 0.022	0.812± 0.017
STRA	15	181	68	50	18	24.038	0.097 ± 0.008	7.563± 1.63	6.831	0.767± 0.025	0.770± 0.022
INSH	13	182	67	49	18	23.91	0.096 ± 0.074	8.125± 1.82	7.066	0.782± 0.022	0.778± 0.025
ABNE	18	173	76	58	18	25.307	0.102 ± 0.006	8.5± 2.48	7.126	0.747± 0.023	0.755± 0.029

743 Na = number alleles, C = conserved sites, v = variable sites, Pi = parsimony informative sites, S = singletons, k= average number of pairwise differences, π = nucleotide
744 diversity
745 Allelic richness based on minimum size of 13 individuals; H_O = observed heterozygosity, H_E = expected heterozygosity
746
747
748
749
750
751
752
753
754
755
756
757
758

Table 2. Analyses of Molecular Variance of MHC and microsatellite data of Scottish highland red deer. Populations were nested according to their location relative to the Great Glen.

MHC alleles

AMOVA analysis	df	SS	MS	Est. var.	% of total var.	Phi	P-value
Among regions	1	0.0937	0.937	0	0	PhiRT = -0.054	0.989
Among populations/regions	2	7.542	3.771	0.134	6	PhiPR = 0.058	0.036
Within population	44	95.417	2.169	2.169	94	PhiPT = 0.007	0.334
Total	47	103.9		2.302	100		

Microsatellites

AMOVA analysis	df	SS	MS	Est. var.	% of total var.	Phi	P-value
Among regions	1	53.502	53.502	0.537	4	PhiRT = 0.036	0.001
Among populations/regions	2	57.041	28.521	0.628	4	PhiPR = 0.043	0.001
Within population	89	1240.9	13.943	13.943	92	PhiPT = 0.077	0.001
Total	92	1351.5		15.108	100		

774 **Table 3.** Summary of MHC allelic variation studies in Cervidae. Please note that studies only assessing variation of MHC DRB expressed loci are not included in this table.
775 Superscript numbers near the species indicate the publication associated with the study.
776

Species	Wild/ Intensively managed/Farmed/Captive	Country	No. populations	No. individuals	No. alleles	Technique
*Cervus elaphus**	Wild	Scotland	4	48	25	Cloning+sequencing (48 clones), 454
C. elaphus[1, 2]	Intensively managed	Spain	1	94	18	SSCP+ cloning (at least 5 independent clones)
C. elaphus[3]	?	Norway	?	20	High	SSCP + direct sequencing homozygotes, cloning-sequencing homozygotes (at least 3 clones)
C. elaphus[3]	?	New Zealand	?	50	49	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
C. elaphus[4]	Wild	Poland	2 (distant)	152	46	Illumina sequencing
Capreolus capreolus[5]	Wild	France	3 (distant)	270	10	454 sequencing
Capreolus capreolus[6]	Managed	France	1	71	4	454 sequencing
C. capreolus[3]	?	Norway	?	40	2	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
C. capreolus[3]	?	Sweden	?	22	4	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
Rangifer tarandus[3]	Wild	Norway	?	20	6	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
R. tarandus[3]	Farmed	Norway	?	20	5	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
Rangifer tarandus[7]	Wild	Canada	5 (distant)	114	19	PCR + Sanger Sequencing
Alces alces[3]	Wild	Sweden	?	198	6	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
Alces alces[3]	Wild	Norway	?	20	7	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)

Dama dama[3]	?	Norway	?	20	2	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
D. dama[3]	?	Sweden	?	30	2	SSCP + direct sequencing homozygotes, cloning-sequencing heterozygotes (at least 3 clones)
Odocoilus virginianus[8]	Intensively manged	USA	1	128	15	SSCP+cloning+ Sanger sequencing
O. virginianus[9]	Intesively managed	USA	1	150	15	SSCP+cloning+ Sanger sequencing
O. virginianus[10]	Wild/Intensively managed	USA	7 (distant)	126	18	SSCP+cloning+ Sanger sequencing
Moschus berezovskii[11]	Captive & wild, museum	China	3 (distant)	20 (captive), 26 (wild),	10	SSCP + cloning of the heterozygotes+ Sanger sequencing
M. berezovskii[12]	Captive	China	3 (distant)	51	17	Cloning+Sanger sequencing (at least 15 clones)
M. berezovskii[13]	Captive	China	1	52	6	Sanger sequencing + cloning of heterozygous (minimum 16 clones)+ Sanger sequencing
Alces alces[14]	wild	Sweden	5 (distant)	30	7	SSCP+ Sanger sequencing
A. alces[14]	Wild	Canada	7 (distant)	19	4	SSCP+ Sanger sequencing

Publications: 1. Fernández de Mera et al. (2009a). 2. Fernández de Mera et al. (2009b). 3. Mikko et al.. (1999). 4. Buczek et al. (2016). 5. Quéméré et al. (2015). 6. Vanpé et al. (2016). 7. Kennedy et al. (2011). 8. Ditchkoff et al. (2005). 9. Ditchkoff et al. (2001). 10. Van Den Bussche et al. (2002). 11. Yao et al. (2015). 12. Xia et al. (2016). 13. Cai et al. (2015). Mikko and Andersson (1995). * Present study.