MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 740–752 doi: 10.1111/j.1365-294X.2010.04975.x

Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*)

K. CAMMEN,*[1] J. I. HOFFMAN,*[2] L. A. KNAPP,† J. HARWOOD‡ and W. AMOS*

*Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK, †Department of Biological Anthropology, University of Cambridge, Pembroke Street, Cambridge CB2 3DZ, UK, ‡Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife KY16 8LB, UK

Abstract

Pathogen-driven balancing selection maintains high genetic diversity in many vertebrates, particularly in the major histocompatibility complex (MHC) immune system gene family, which is often associated with disease susceptibility. In large natural populations where subpopulations face different pathogen pressures, the MHC should show greater genetic differentiation within a species than neutral markers. We examined genetic diversity at the MHC-DQB locus and nine putatively neutral microsatellite markers in grey seals (*Halichoerus grypus*) from eight United Kingdom (UK) colonies, the Færoe Islands and Sable Island, Canada. Five DQB alleles were identified in grey seals, which varied in prevalence across the grey seal range. Among the seal colonies, significant differences in DQB allele and haplotype frequencies and in average DQB heterozygosity were observed. Additionally, the DQB gene exhibited greater differentiation among colonies compared with neutral markers, yet a weaker pattern of isolation by distance (IBD). After correcting for the underlying IBD pattern, subpopulations breeding in similar habitats were more similar to one another in DQB allele frequencies than populations breeding in different habitats, but the same did not hold true for microsatellites, suggesting that habitat-specific pathogen pressure influences MHC evolution. Overall, the data are consistent with selection at MHC-DQB loci in grey seals with both varying selective pressures and geographic population structure appearing to influence the DQB genetic composition of breeding colonies.

Keywords: grey seal, isolation by distance, major histocompatibility complex, microsatellites, selection

Received 27 May 2010; revision received 5 November 2010; accepted 23 November 2010

Introduction

The major histocompatibility complex (MHC) is a family of highly polymorphic genes involved in antigen presentation within the vertebrate immune system (Klein & Sato 2000a,b). Most of the variation within the MHC is centred on the peptide-binding region, the part of the molecule that recognizes and binds antigens (Hughes & Nei 1988, 1989). It appears that populations carrying high MHC diversity are often capable of responding to a wider range of pathogens (Sommer 2005). Moreover, as an individual can only possess two alleles at any one locus, gene duplication is frequent and, as a consequence, many species express multiple types of MHC molecules to maximize the range of pathogens that their immune system can recognize (Doherty & Zinkernagel 1975).

Since the discovery of the MHC, diversity at these genes and resistance to disease have been assumed to be intimately linked. Many studies, conducted in a wide

Correspondence: William Amos, Fax: +44 1223 336676; E-mail: w.amos@cam.ac.uk

[1]Present address: Duke University Marine Laboratory, 135 Duke Marine Lab Road, Beaufort, NC 28516, USA.

[2]Present address: Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany.

range of species, lend empirical support to this assumption. When a single disease is studied, the usual pattern reported is one where a single specific MHC allele is associated with some level of resistance (Hill *et al.* 1991; Grimholt *et al.* 2003; McClelland *et al.* 2003). When multiple pathogens are involved, there is instead often a correlation between MHC heterozygosity and disease resistance (Penn *et al.* 2002; McClelland *et al.* 2003; Wegner *et al.* 2003). The latter observations suggest heterozygote advantage, one form of balancing selection. However, heterozygote advantage alone struggles to maintain more than a few alleles in a population (Kimura & Crow 1964). MHC allelic diversity may instead be maintained by cyclically fluctuating selection pressure due to pathogens co-evolving to resist the most common host MHC alleles, i.e. rare-allele advantage (frequency-dependent selection), another form of balancing selection (Slade & McCallum 1992). A further possibility is that selection pressure varies across time or space (Hedrick 2002) or that some combination of these three processes is involved in maintaining MHC allele diversity (Spurgin & Richardson 2010).

Attempts to clarify the exact relationship between MHC diversity and disease susceptibility are often confounded by the effects of demographic history. One commonly used approach is the comparison of MHC diversity and levels of genetic diversity of presumed neutral markers. Unfortunately, many of these studies have focused on either bottlenecked or fragmented populations, where the impact of drift tends to be unusually large and may overwhelm anything but strong selection (Landry & Bernatchez 2001; Aguilar & Garza 2006; Biedrzycka & Radwan 2008), thereby reducing the signal. In most of these cases, congruence between neutral markers and the MHC has been reported. In a few studies, MHC diversity has been compared with neutral genetic diversity in large populations and significant differences in population differentiation at the two marker types have been observed. However, neither marker type shows consistently stronger population differentiation across studies. Depending on the study system, population differentiation at the MHC may be stronger because of diversifying selection (Miller *et al.* 2001; Ekblom *et al.* 2007; Loiseau *et al.* 2009) or weaker because of purifying selection (Fraser *et al.* 2010a). In some cases, associations between MHC genotypes and disease have been sought (Miller *et al.* 2004; Fraser & Neff 2010); though, few studies attempt to distinguish between different models of MHC evolution.

A potential difference between the expectations of heterozygote advantage and allele-specific responses is the way in which MHC diversity correlates with habitat type. If diversity is maintained by heterozygote advantage, individuals should carry similar alleles across a species' range. In contrast, under allele-specific selection, populations exposed to similar pathogens because they live in similar habitats should appear more similar to each other than expected by chance, even if they are geographically distant. Equally, animals living in different but adjacent habitats may possess dissimilar MHC alleles. An interesting example of this involves the northern elephant seal (*Mirounga angustirostris*), where it was shown that the MHC-DQA differed significantly between two breeding colonies, but mitochondrial DNA and microsatellites showed no significant differences in allele frequency between populations (Weber *et al.* 2004). Similarly, California sea lions (*Zalophus californianus*) exhibited significant MHC-DRB allele frequency differences between colonies in close proximity, although in this case data from neutral markers were not presented (Bowen *et al.* 2006).

These two examples emphasize the utility of pinnipeds for exploring MHC evolution. Most pinniped species breed in colonies and show strong natal site fidelity. Within these colonies, individuals may be exposed to unusually high levels of pathogens owing to both the close proximity of other individuals and the accumulation of faecal material. Moreover, while some species breed in rather uniform habitats, others use a variety of substrates, creating conditions under which colonies may differ markedly in the range of pathogens to which individuals are exposed. The grey seal (*Halichoerus grypus*) is one such species, breeding in many partially isolated colonies (Pomeroy *et al.* 2000) that vary considerably in habitat type, including rocky, sandy, muddy and shingle substrates. Bacterial infections cause high mortality in grey seal pups at most breeding colonies, but mortality rates differ among colonies (Anderson *et al.* 1979; Baker 1984; Baker & Baker 1988). These differences in mortality appear to be related to colony topography and substrate. In general, colonies where seals breed inland on muddy substrate have higher rates of infection-related mortality (Baker 1984; Baker & Baker 1988) than colonies on sandy beaches (Baker 1984). We would therefore expect that class II MHC molecules, which respond to bacterial infections (Klein & Sato 2000a), may be subject to colony-specific selection.

In this study, we analysed MHC class II DQB variation among ten grey seal breeding colonies in the North Atlantic. Previous studies have shown genetic differentiation at presumably neutral microsatellites between two of the UK breeding colonies (Allen *et al.* 1995). We expanded this microsatellite analysis to include additional colonies and used these neutral markers to estimate a baseline of differentiation among colonies against which to compare the MHC.

Materials and methods

Study sites and tissue sampling

Tissue samples were collected between 1991 and 2003 from 1132 live grey seals at 10 breeding colonies (Table 1). The sampling protocol followed Bean *et al.* (2004). Skin samples were stored individually in 20% dimethyl sulphoxide saturated with salt (Amos & Hoelzel 1991) and kept at −20 °C. The study focused primarily on the Northeast Atlantic grey seal population (Fig. 1). Sable Island, Canada, was included as an outgroup representing the geographically isolated Northwest Atlantic population.

The 10 colonies represented four different types of habitat in which grey seals breed (Table 1). For our analyses, each of the colonies was categorized as one of the following: sandy beach, shingle beach, rocky beach or inland. Inland sites were often located near pools of water and characterized by muddy substrate. As a

Table 1 Environmental and demographic characteristics of the grey seal breeding colonies included in the study

Breeding colony	Habitat	Pup production	Collection year(s)	No. of MHC genotypes	No. of microsatellite genotypes
Orkneys (58°41′N, 3°7′W)		19 300*			
Corn Holm	Sand		1995	40	75
Stroma	Shingle		1995 and 1999	46	154
Muckle Greenholm	Inland		1995	45	159
Monachs (57°31′N, 7°40′W)	Sand	11 600*	1994	48	94
North Rona (59°7′N, 5°49′W)	Inland	1000*	1996	50	229
Isle of May (56°11′N, 2°34′W)		2500*			
Pilgrims Haven	Shingle		1991 and 1992	49	99
Inland site	Inland		1999 and 2002	65	132
Farnes (55°37′N, 1°37′W)	Rocky	1250*	1997	37	94
Færoes (62°00′N, 6°47′W)	Rocky	NA†	1999	48	68
Sable Island (43°57′N, 59°55′W)	Sand	25 000‡	2003	28	28
Total				456	1132

*Pup production estimates from 2006 (SCOS 2007).
†No pup production estimate is available for the Færoese grey seal population. However, Mikkelsen (2007) estimates a total population of 1000–2000 individuals.
‡Pup production estimate from 1997 (Bowen *et al.* 2003).

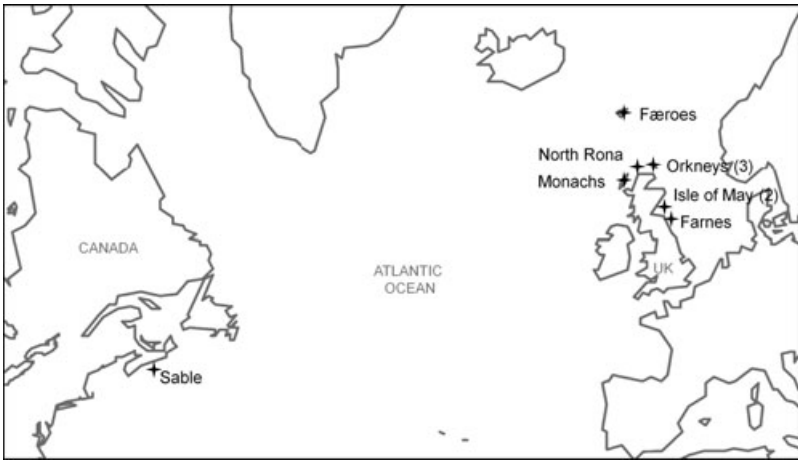

Fig. 1 Grey seal breeding colonies (+) sampled in the North Atlantic. Two sites were sampled from the Isle of May: Pilgrims Haven and an inland site. Three sites were sampled from the Orkneys: Corn Holm, Stroma and Muckle Greenholm.

measure of habitat similarity, we used a simple binary coding of 1 = same, 0 = different and referred to this metric as 'habitat distance'.

DNA extraction and genotyping

Genomic DNA was extracted from tissue samples using standard phenol–chloroform procedures (Sambrook *et al.* 1989). Data for nine microsatellite loci (Table 2) were already available for some samples collected up to and including 1999 (Worthington Wilmer *et al.* 1999; Gaggiotti *et al.* 2002). The remaining samples were genotyped at the same panel of markers as described by Allen *et al.* (1995). Samples from 456 seals were genotyped at the DQB exon 2 region using single-strand conformation polymorphism (SSCP) gel electrophoresis. Briefly, DQB PCR amplification was conducted in a final volume of 12 μL, containing 2 μL of DNA, 1× Buffer A (10 mM Tris–HCl pH 8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Tween, 0.1% Gelatin, 0.1% IGEPAL), 2.5 mM $MgCl_2$, 0.3 μM of each primer, 0.1 mM dATP, dGTP, dTTP, 0.01 mM dCTP, 0.005 units of *Taq* polymerase and 0.01 μCi [α^{33}P]-dCTP. The primers followed Bean (2005): forward 5′-TCGTGTACCAGTTTAAGGGC; reverse 5′-CGCTCACCTCGCCGCTC. The PCR profile included an initial denaturation at 94 °C for 4 min, followed by 35 cycles of 94 °C for 1 min, 65 °C for 30 s and 72 °C for 30 s, followed by a final extension at 72 °C for 5 min. Prior to electrophoresis, 10 μL of stop solution (80% formamide, 10 mM EDTA, 1 mg/mL xylene cyanol and 1 mg/mL bromophenol blue) was added to DQB PCR products, which were then denatured for 5 min at 96 °C and put on ice. Alleles were separated on a 6% polyacrylamide gel (Sequagel, National Diagnostics) with 0.05% glycerol run at approximately 4 °C and 15 watts for 7.5 h. SSCP patterns were visualized using autoradiography. Samples were run alongside a set of single-allele clones to standardize scoring. Distinct SSCP banding patterns are referred to as alleles, although they cannot be assigned with confidence to a particular locus because of co-amplification of products from multiple genes. Following previous usage, we refer to SSCP multi-allele banding patterns as haplotypes (see Brown *et al.* 1993).

Cloning and sequencing

The DQB exon 2 was cloned and sequenced from 17 individuals, representing all 10 breeding colonies and all of the observed SSCP bands. Amplified fragments were gel-isolated (QIAquick Gel Extraction kit; Qiagen), ligated into pGem®-T Easy Vectors (Promega) and transformed into Subcloning Efficiency™ DH5α™ Chemically Competent Cells (Promega). Following growth and blue-white selection, DNA was isolated from bacterial colonies by boiling lysis (Sambrook *et al.* 1989) and genotyped using SSCP. A subset of clones including multiple copies of each observed banding pattern was then sequenced in both directions (Alpha Biolabs). As MHC gene sequencing can be prone to artefacts generated by *Taq* errors and heteroduplex mismatch repair (van Oosterhout *et al.* 2006), only DQB sequences that were observed in multiple clones from at least two individuals were confirmed.

Data analyses

Grey seal DQB sequences were aligned and edited using BioEdit 7.0 (Hall 1999). As SSCP cannot distinguish between homozygotes and heterozygotes, DQB allele frequencies were estimated using the frequency of individuals in a population that did not express the allele, for example $\text{freq}_A = 1-\sqrt{\text{freq}_{NotA}}$. This equation assumes Hardy–Weinberg equilibrium, which was tested by comparing SSCP-estimated ('observed') allele frequencies with those expected from the haplotype frequencies for each colony using chi-squared tests. Here and elsewhere, tests were repeated both including and excluding Sable Island because of this colony's marked geographic isolation. Only haplotypes for which all expected values exceeded one were included in the chi-squared tests, and all resulting *P* values were subject to sequential Bonferroni correction to compensate for multiple tests (Rice 1989). DQB heterozygosity, calculated as the average number of alleles per individual, was compared among colonies using a one-way ANOVA.

Microsatellite loci were tested for deviations from Hardy–Weinberg equilibrium and for linkage disequilibrium between pairs of loci at each population using Arlequin 3.1 (Excoffier *et al.* 2005). Outlier microsatellite markers that may be linked to genes under positive selection were identified using the Lositan Selection Workbench (Beaumont & Nichols 1996; Antao *et al.* 2008), which evaluates the relationship between F_{ST} and heterozygosity and identifies deviations from neutral expectations under an island model of migration. The Lositan model was run with 10 000 simulations, assuming a stepwise mutation model and a forced 'neutral' mean F_{ST}.

As SSCP genotypes do not conform to the standard diploid format, F_{ST} could not be calculated in the standard way. Instead, genetic differentiation was calculated for both microsatellites and the MHC using two methods manually implemented in Microsoft Excel. Theta (Θ), an estimate of genetic differentiation that corrects for differences in sample size among populations, was calculated as described by Weir & Cockerham (1984). G'_{ST}, an estimate of genetic differentiation that

Table 2 Summary of the microsatellite loci used in this study, including literature sources and number of alleles per locus, observed heterozygosity (H_O), and results of an exact test for Hardy–Weinberg equilibrium (*P* values without Bonferroni correction)

	Isolated from sp.	References	No. of alleles	Sable		Faeroes		Monachs		North Rona		Stroma		Corn Holm		Muckle Greenholm		Isle of May		Pilgrims Haven		Farnes	
				H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*	H_O	*P*
Hg3.6	*Halichoerus grypus*	Allen *et al.* (1995)	8	0.82	0.79	0.85	0.34	0.74	0.07	0.78	0.78	0.79	0.08	0.80	0.63	0.72	0.04	0.78	0.76	0.77	0.44	0.90	0.22
Hg4.2	*H. grypus*	Allen *et al.* (1995)	10	0.71	0.05	0.81	0.86	0.63	0.56	0.63	0.55	0.62	0.20	0.61	0.36	0.63	0.66	0.73	0.08	0.72	0.76	0.70	0.32
Hg6.1	*H. grypus*	Allen *et al.* (1995)	6	0.62	0.85	0.56	0.42	0.58	0.28	0.62	0.18	0.64	0.94	0.60	0.38	0.64	0.31	0.63	0.99	0.69	0.74	0.57	0.10
Hg6.3	*H. grypus*	Allen *et al.* (1995)	6	0.82	0.91	0.85	0.14	0.74	0.34	0.74	0.11	0.72	0.43	0.84	0.09	0.75	0.13	0.71	0.28	0.76	0.44	0.70	0.17
Hg8.9	*H. grypus*	Allen *et al.* (1995)	13	0.61	0.07	0.68	0.07	0.79	0.61	0.86	0.18	0.81	0.02	0.83	0.60	0.79	0.39	0.82	0.83	0.74	0.02	0.87	0.15
Hg8.10	*H. grypus*	Allen *et al.* (1995)	11	0.67	0.10	0.78	0.31	0.80	0.48	0.82	0.78	0.81	0.36	0.81	0.66	0.77	0.24	0.76	0.77	0.78	0.16	0.83	0.22
Hgd.2	*H. grypus*	Allen *et al.* (1995)	8	0.74	0.28	0.66	0.34	0.84	0.02	0.68	0.29	0.59	0.13	0.59	**0.00**	0.68	0.55	0.75	0.35	0.74	0.72	0.71	0.48
Pv9	*Phoca vitulina*	Goodman (1997)	8	0.64	0.45	0.86	0.11	0.81	0.84	0.79	0.19	0.76	0.45	0.75	0.34	0.80	0.24	0.81	0.46	0.82	0.18	0.76	0.19
Pv11	*P. vitulina*	Goodman (1997)	9	0.64	0.66	0.68	0.03	0.73	0.16	0.71	0.02	0.69	0.07	0.70	0.11	0.70	0.48	0.71	0.89	0.67	0.68	0.71	0.85

P values that are shown in bold remained significant following table-wide sequential Bonferroni correction for multiple tests with $\alpha = 0.05$.

corrects for differences in heterozygosity among loci, was calculated as described by Hedrick (2005) and Nei (1973). For both analyses, the MHC was considered one locus with four alleles. Weir & Cockerham's (1984) methods require as input the number of populations sampled (r), the size of each sample (n), the frequency of each allele (p) and the proportion of heterozygous individuals (h). As SSCP cannot distinguish between homozygotes and heterozygotes, the proportion of heterozygotes was calculated for each locus from single-allele frequencies, assuming the populations were in Hardy–Weinberg equilibrium. This assumption was tested as described earlier. The standard errors in the overall Θ and G'_{ST} reported for the microsatellite loci were estimated by jackknifing over loci. Our estimates of Θ were compared to values for Θ from FSTAT 2.9.3 (Goudet 2001), which calculates Weir & Cockerham's (1984) Θ with 15 000 bootstrap replicates. There was no significant difference between our overall estimate of microsatellite Θ (0.009, SE = 0.002) and FSTAT's overall estimate of microsatellite Θ (0.009, SE = 0.002).

Genetic differentiation between pairs of populations was estimated using Θ. Pairwise Θ values were regressed against geographic distances between each pair of populations to yield an isolation-by-distance (IBD) plot. Geographic distances were estimated as the shortest path that could be drawn between colonies using a straight line and avoiding land barriers. To test for IBD, we used the contributed package 'ecodist' in R (R Development Core Team 2008), which tests for an association between two distance matrices using a straight Mantel test for an association between one distance matrix and two or more predictor matrices using partial Mantel tests. However, we were also interested in testing models of the form genetic distance predicted by geographic distance, habitat distance, and the interaction between habitat and geographic distance. As 'ecodist' does not allow interaction terms, we wrote our own R-script to conduct a variant of the Mantel test. First, the desired general linear model was fitted (genetic distance $\sim$ geographic distance + habitat distance + habitat distance $\times$ geographic distance). Next, significance was tested by repeatedly scrambling habitat type across the 10 sampling sites, regenerating the habitat distance matrix and then refitting the model with these scrambled data. Significance was tested by counting the number of times the model based on scrambled habitat data yielded a lower AIC value compared to the original model. No model simplification was implemented during this procedure because the IBD component is unaffected by randomizing habitat, so the scrambling should only reveal a significant effect if either habitat or habitat–geography interaction improves the model.

Results

Sequence analysis

Five DQB sequences were identified in grey seals and named *Hagr-DQB*01-05* following the convention of Klein *et al.* (1990). The five sequences were all highly variable, differing by up to 13.7% in nucleotide sequence (Fig. 2) and 25.8% in amino acid sequence (Fig. 3). MHC primers often amplify multiple loci because of the similarity in sequences between duplicated loci within the MHC (Bowen *et al.* 2002; Baker *et al.* 2006), and in this case, individual grey seals possessed up to four *Hagr*-DQB sequences, indicating at least two DQB loci. The grey seal DQB sequences were closely related to the human MHC gene HLA-DQB1 (83.4% homology with *Hagr-DQB*01*) and the dog MHC gene DLA-DQB1 (86.0% homology with *Hagr-DQB*01*).

Geographic variation in the MHC

SSCP was only able to distinguish four patterns from the five *Hagr*-DQB sequences, referred to as alleles A (*Hagr-DQB*01* and **04*), B (**03*), C (**02*) and D (**05*). Single-sequence clones for *Hagr-DQB*01* and **04* produced SSCP banding patterns that were too similar to be confidently differentiated across individuals, and therefore the two sequences were scored as one SSCP allele. Across all colonies, allele frequencies estimated from observed SSCP patterns correlated strongly with expected allele frequencies calculated from haplotype frequencies ($r^2 = 0.996$), supporting the assumption of a diploid system in Hardy–Weinberg equilibrium.

Significant differences were observed in both allele and haplotype frequencies among the ten colonies (Table 3). Allele frequency differences were significant for allele C ($\chi^2 = 33.09$, d.f. = 9, $P < 0.001$) and allele D ($\chi^2 = 20.80$, d.f. = 9, $P < 0.05$). Thirteen haplotypes were observed, and haplotype frequency differences were significant for haplotype ACD ($\chi^2 = 26.76$, d.f. = 9, $P < 0.01$). When Sable Island was removed from the analysis, frequency differences became not significant for alleles C ($\chi^2 = 17.75$, d.f. = 8, $P = 0.023$, not significant following sequential Bonferroni correction) and D ($\chi^2 = 15.42$, d.f. = 8, $P = 0.051$), but remained significant for haplotype ACD ($\chi^2 = 23.70$, d.f. = 8, $P < 0.05$). Analysing all alleles together revealed significant heterogeneity ($\chi^2 = 58.56$, d.f. = 27, $P < 0.001$); though, significance was lost when Sable Island was removed ($\chi^2 = 34.36$, d.f. = 24, $P = 0.078$). A significant difference was also observed among the ten populations in DQB heterozygosity (one-way ANOVA: $F_{9,446} = 2.97$, $P < 0.01$). This difference remained significant even after excluding Sable Island from the analysis ($F_{8,419} = 2.85$, $P < 0.01$).

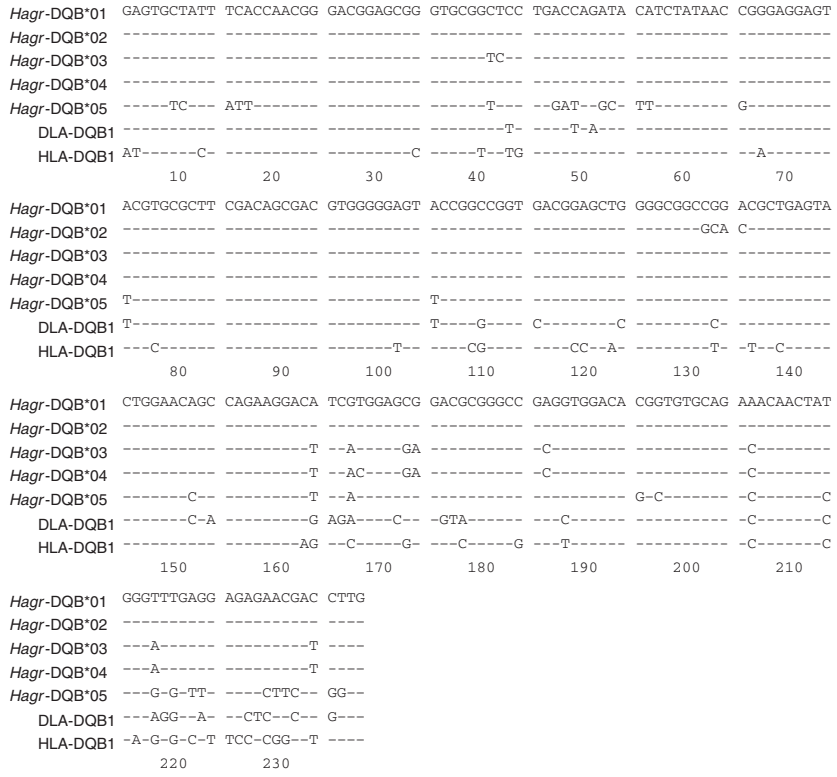

Fig. 2 Nucleotide sequence alignment of the five grey seal sequences (*Hagr-DQB*01-05*, GenBank Accession nos. HQ456122–HQ456126) and two outgroups, human (HLA-DQB reference sequence, NM_002123) and dog (DLA-DQB1 reference sequence, AF043908) sequences from the DQB exon 2 region. A dash indicates nucleotide identity with the consensus sequence.

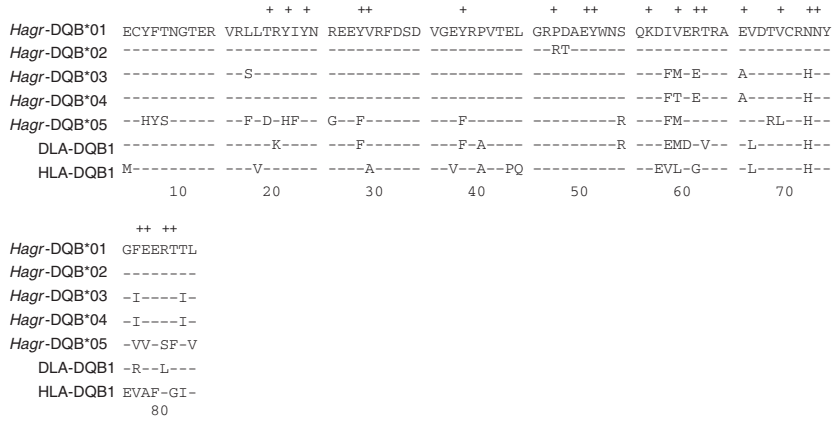

Fig. 3 Predicted protein sequence alignment of the five grey seal DQB sequences (*Hagr-DQB*01-05*) and two outgroups, human and dog. A dash indicates amino acid identity with the consensus sequence. A plus sign (+) above an amino acid indicates a putative peptide-binding region site (Brown *et al.* 1993).

Significant differences in allele and haplotype frequencies were also observed among habitat types (Table 3). Based on all ten colonies, allele D was observed more frequently than expected in sandy breeding habitats and less frequently than expected in inland breeding habitats ($\chi^2 = 15.72$, d.f. = 3, $P < 0.01$). Removing Sable Island reduced but did not remove significance for allele D ($\chi^2 = 10.63$, d.f. = 3, $P < 0.05$) and uncovered stronger heterogeneity for allele C ($\chi^2 = 17.80$, d.f. = 3, $P < 0.001$), which was in excess at sandy habitats and in deficit at shingle habitats. Without Sable Island, the frequency of haplotype ACD also

Table 3 DQB allele frequency data for each population and each habitat type

Habitat type	Sample size	Allele frequency			
		A	B	C	D
Inland	160	0.63	0.10	0.27	0.07
Muckle Greenholm	45	0.58	0.11	0.29	0.08
North Rona	50	0.63	0.12	0.32	0.07
Isle of May	65	0.67	0.07	0.22	0.06
Rocky	85	0.66	0.13	0.32	0.12
Færoes	48	0.62	0.13	0.37	0.13
Farnes	37	0.72	0.11	0.25	0.10
Sand	116	0.65	0.08	0.31	0.18
Corn Holm	40	0.58	0.12	0.37	0.11
Monachs	48	0.62	0.05	0.42	0.21
Sable	28	0.81	0.06	0.02	0.22
Shingle	95	0.71	0.08	0.21	0.09
Stroma	46	0.71	0.08	0.22	0.12
Pilgrims Haven	49	0.71	0.09	0.19	0.07

differed among habitat types ($\chi^2 = 15.18$, d.f. = 3, $P < 0.01$), being more frequent at sandy habitats. Analysing all alleles together revealed significant heterogeneity among habitats, regardless of whether Sable Island was included ($\chi^2 = 19.03$, d.f. = 9, $P < 0.001$) or excluded ($\chi^2 = 20.98$, d.f. = 9, $P < 0.001$). DQB heterozygosity also varied significantly among habitats, again regardless of inclusion of Sable Island (including Sable: $F_{3,448} = 4.05$, $P < 0.01$; excluding Sable: $F_{3,424} = 5.41$, $P < 0.01$).

Comparison between MHC and neutral markers

All microsatellite loci were in Hardy–Weinberg equilibrium at each of the breeding colonies, except in one case, Hgd.2 at Corn Holm (Table 2). Linkage disequilibrium tests resulted in significant P values for only five pairs of microsatellite loci (data not shown), but when testing 36 pairs of loci across 10 populations, it is expected that up to 18 pairs show significant linkage disequilibrium ($P < 0.05$) by chance. Overall, the data suggested the absence of linkage disequilibrium. The genetic differentiation among populations was greater in magnitude at the DQB than at the nine microsatellite loci for both Θ and G'_{ST} (Table 4). This difference remained when Sable Island was excluded from the analysis, although the genetic differentiation among populations at both marker types decreased. In both cases, DQB Θ was above two standard errors calculated around the microsatellite Θ. Although the two estimates of genetic differentiation among populations, Θ and G'_{ST}, returned different absolute values, the trends observed in differences among populations and differences among loci were similar. The fact that two separate estimates showed similar trends indicated that the differences observed between the MHC and microsatellites markers were robust both to differences in sample size among populations and to differences in heterozygosity among markers, a concern when comparing loci with different mutation rates (Hedrick 1999).

Table 4 Estimates of genetic differentiation among colonies. Θ was calculated as described by Weir & Cockerham (1984). G'_{ST} was calculated as described by Hedrick (2005) and Nei (1973). Standard errors were calculated by jackknifing over microsatellite loci

	Θ	Θ SE	G'_{ST}	G'_{ST} SE
All populations				
MHC	0.020		0.101	
Microsatellites	0.009	0.002	0.068	0.019
Microsatellites except Hg6.3	0.007	0.001	0.059	0.020
Without Sable Island				
MHC	0.010		0.086	
Microsatellites	0.005	0.001	0.039	0.011
Microsatellites except Hg6.3	0.004	0.001	0.033	0.010

Although microsatellites are presumed neutral, there is increasing evidence both in the grey seal (Bean *et al.* 2004) and in other species (e.g. Antarctic fur seals, Hoffman *et al.* 2004) that some loci may be individually associated with key fitness traits, implying linkage to genes under selection. Consequently, Lositan was used to identify F_{ST} outliers that may be candidates of selection. One microsatellite locus, Hg6.3, was identified as an F_{ST} outlier ($P < 0.01$, Table 5); though, this probability was no longer significant when Sable Island was removed from the analysis ($P = 0.183$). Hg6.3 heterozygosity is associated with both pup survival in this species (Bean *et al.* 2004) and body size in male fur seals (Hoffman *et al.* 2010). When this locus was excluded from the analysis, the difference between genetic differentiation among populations at the DQB and at the remaining microsatellite markers became greater (Table 4).

We next tested for IBD patterns using Mantel tests. Both DQB and overall microsatellite Θ exhibited statistically significant regressions, regardless of whether Sable Island was included or excluded (DQB including Sable: Mantel $r = 0.936$, $n = 10$, $P < 0.01$; DQB excluding Sable: Mantel $r = 0.441$, $n = 9$, $P < 0.05$; microsatellites including Sable: Mantel $r = 0.997$, $n = 10$, $P < 0.01$; microsatellites excluding Sable: Mantel $r = 0.737$, $n = 9$, $P < 0.01$). Most individual microsatellite loci also exhibited statistically significant regressions when Sable Island was included, but the regressions for individual microsatellites were more variable and often not significant when Sable Island was removed from the analysis

Table 5 Results of the Lositan analysis (Beaumont and Nichols 1996; Antao *et al.* 2008), showing geographic differentiation (F_{ST}) and expected heterozygosity (H_E) for each of the microsatellite loci, both including and excluding Sable Island

	All colonies			Excluding Sable Island		
	H_E	F_{ST}	P	H_E	F_{ST}	P
Hg3.6	0.803	0.013	0.529	0.801	0.007	0.603
Hg4.2	0.678	0.016	0.660	0.669	0.008	0.674
Hg6.1	0.647	0.020	0.743	0.630	0.002	0.430
Hg6.3	0.771	0.036	**0.991**	0.756	0.012	0.817
Hg8.9	0.811	0.011	0.450	0.816	0.007	0.596
Hg8.10	0.791	0.015	0.627	0.793	0.005	0.497
Hgd.2	0.726	0.014	0.590	0.716	0.007	0.633
Pv9	0.799	0.017	0.713	0.805	0.003	0.379
Pv11	0.693	0.005	0.232	0.698	0.004	0.481

P values represent the probability that a locus is an F_{ST} outlier and a candidate for positive selection. Significant P values ($P > 0.95$) are shown in bold.

(data not shown). Finally, we tested whether genetic distance was influenced by habitat similarity by fitting models in which genetic distance is the response variable and predictors are geographic distance, habitat distance and the interaction between habitat and geographic distance. Significance was tested by repeatedly scrambling the habitat similarity matrix (see Methods for details). Regardless of whether Sable Island was included or excluded, microsatellite distance was unaffected by habitat similarity while genetic distance based on MHC was significantly affected by habitat (DQB including Sable: $P < 0.05$; DQB excluding Sable: $P < 0.05$; microsatellites including Sable: $P = 0.996$; microsatellites excluding Sable: $P = 0.083$). In the MHC model, the interaction term (habitat distance × geographic distance) was significant ($P < 0.01$), indicating that the effect of habitat varies with geographic distance. As the MHC was treated as a single locus while we had nine microsatellites, it was possible that individual microsatellites would show significant associations with habitat even if their average did not. We therefore tested each locus separately, but still found no significant associations (lowest P value = 0.09).

Discussion

This study characterizes the DQB exon 2 region of the MHC in grey seals and investigates diversity at this locus among breeding colonies in the North Atlantic. We find that the MHC, which plays an important role in disease susceptibility (Sommer 2005), exhibits greater differentiation among breeding colonies compared with neutral markers, yet a weaker pattern of IBD. Both observations are consistent with the action of natural selection. Moreover, seal populations that breed in similar habitats are more similar to one another in their MHC genetic composition than seal populations that breed in different habitats, but this is not the case for the microsatellite loci. This suggests that habitat-specific pathogen pressure influences MHC evolution.

MHC genotyping is seldom straightforward. We chose to use SSCP, which is relatively efficient for genotyping large numbers of animals. However, SSCP could only reliably resolve four allelic patterns from the five DQB sequences we identified in grey seals, indicating that one allelic pattern was composed of two different sequences. Moreover, as multiple loci were co-amplified, we were unable to distinguish homozygotes from heterozygotes. However, we were able to show that the frequencies of observed SSCP patterns matched closely those expected based on individual allele frequency estimates under the assumption of Hardy–Weinberg equilibrium. Additionally, a subset of individuals that were genotyped twice to check for errors in scoring confirmed that our genotyping was 100% reliable. We were unable to test for gene expression directly, raising the possibility that some or all of the PCR products that we amplified came from nonexpressed genes. Consequently, any evidence we uncovered of the action of natural selection could potentially reflect selection at closely linked functional genes. It is also possible that spurious trends might arise from our inability to genotype these loci separately. However, we feel this is unlikely because allele frequencies were estimated from the probability of not carrying an allele, which is easy to score and should be robust to anything but large deviations from Hardy–Weinberg equilibrium.

Like other pinnipeds, grey seals have relatively few MHC-DQB alleles (Hoelzel *et al.* 1999; Bowen *et al.* 2002; Lento *et al.* 2003; Weber *et al.* 2004). However, this does not necessarily indicate a lack of selection. Low genetic diversity may be observed because the marine environment presents a more restricted and consistent range of challenges than the terrestrial environment (Slade 1992; but see McCallum *et al.* 2004). Indeed, the greater genetic differentiation observed at the MHC compared with microsatellite markers implies the action of natural selection because the microsatellite data should behave neutrally. Moreover, overall, microsatellites show a stronger pattern of IBD when compared to the MHC; though, individual microsatellite loci are quite variable. This supports the conclusion that selection plays a role in determining the distribution of MHC variability because, with higher overall genetic differentiation, if the MHC were behaving neutrally, the expectation would be that it should also reveal a stronger IBD pattern. Direct comparisons of population

structure between MHC genes and microsatellites may be complicated by the different ways in which these markers evolve. However, we have tried to account for these differences by calculating a standardized measure of genetic differentiation, G'_{ST}, which is independent of genetic variation within populations and allows the fair comparison of loci with different mutation rates (Hedrick 2005).

Selection on the MHC may vary both temporally and/or geographically. MHC allele frequencies vary over time in areas where pathogen pressures also change yearly (Westerdahl *et al.* 2004; Charbonnel & Pemberton 2005; Fraser *et al.* 2010b). Unfortunately, our data set did not allow us to test for temporal changes, but, despite a paucity of data, it seems to us unlikely that any temporal differences would be greater than the differences between contrasting breeding substrates. For example, previous research showed that pup mortality at North Rona varied little between 1959 and 1968 (Boyd & Campbell 1971) or between 1980, 1985 and 1986 (Baker 1984, 1988). The latter study (Baker 1988) did find a difference in the pathogens present among years, indicating the potential for pathogen pressure to change over time, though by no means proving a direct link to selection at the MHC. Further research in this area is clearly desirable.

The importance of geographic variation in selection pressure at the MHC mediated by differences in breeding substrate is supported both by our current results and by previous work. Several studies on pup mortality at grey seal breeding colonies have found an association between island topography and pup mortality, in particular the rate of mortality by infection, which is greatest on inland breeding substrates and lowest on sandy substrates (Anderson *et al.* 1979; Baker 1984; Baker & Baker 1988). The infective agents, primarily opportunistic bacteria, have also been found to vary across colonies (Anderson *et al.* 1979; Baker 1988). With this knowledge from previous work, one might reasonably expect that seals breeding at different sites would be exposed to different pathogens, leading to a pattern of MHC genetics of the sort we report. In other words, as we observed, seal populations breeding in similar habitats would be expected to be more similar in their MHC allele frequencies than expected based on geographic proximity alone. The significant differences we observed in MHC allele frequencies were most apparent between colonies characterized by sandy substrates and inland substrates, as would be expected based on known differences in infection rates between these habitat types.

Geographic variation in MHC allele frequencies among populations of a single species has been observed in other systems, which also exhibited greater genetic differentiation at the MHC than at neutral microsatellite markers (Miller *et al.* 2001; Ekblom *et al.* 2007; Loiseau *et al.* 2009). This geographic variation is often explained as we propose here, in terms of variation in pathogen pressures among populations leading to geographically localized adaptations of the MHC and varying levels of balancing selection among populations (Miller *et al.* 2001). A small number of studies have found evidence of purifying or directional selection at the MHC across populations, usually explained by a pervasive pathogen present across the sampled range (i.e. Fraser *et al.* 2010a). It is obvious from our results and from previous studies of grey seal mortality among colonies (Anderson *et al.* 1979; Baker 1988) that grey seals are not exposed to any one pathogen consistently across their range in the Northeast Atlantic.

While the differences that we observed in MHC allele frequencies among grey seal colonies were significant, they were not large in magnitude (Table 3). Similarly, low but significant differences in MHC allele frequency have been reported elsewhere (Ekblom *et al.* 2007), but other studies have found much greater differences in allele frequencies among populations (Miller *et al.* 2001; Bowen *et al.* 2006). The magnitude of allele frequency differences may reflect the strength of selection on the MHC, the degree of variation in selection pressures among colonies, or, simply, high levels of gene flow between colonies in this highly mobile species. Further studies of pathogen presence and virulence across grey seal colonies would be required to distinguish between these explanations for the relatively low observed differences in allele frequency.

In the model of how MHC allele frequencies vary with habitat and geographic distance, the interaction term is significant. This might be expected given the large range of distances over which seals habitually move. At very short distances, high levels of gene flow may well prevent appreciable differences in MHC allele frequencies from developing. Similarly, at very large distances (across the Atlantic), gene flow is negligible, and IBD may dominate. Indeed, even in similar habitats separated by great distances, the pathogens present may have diverged. Plausibly, therefore, a strong influence of habitat type is only likely to be found between colonies linked by little but not zero gene flow. The power of our analysis would inevitably be improved by increasing the number of colonies sampled, particularly over smaller spatial scales. Additionally, to understand further the mechanism of natural selection in this system, samples should be collected from pups that died of infection in different habitats to test the prediction that such animals will differ from the colony average in the MHC alleles they carry.

Acknowledgements

We thank Kyi Bean, Don Bowen, Ailsa Hall, Rob Harris, Bernie McConnell, Penny Allen, Paddy Pomeroy, Sean Twiss and Dorete Bloch for collecting samples. We also thank Karina Acevedo-Whitehouse, Nicholas Mundy and two anonymous reviewers for their comments on an earlier version of this paper. KC was funded by the Cambridge University Overseas Trust, and JH was supported by a British Antarctic Survey Strategic Alliance Fellowship.

References

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

Allen P, Amos W, Pomeroy P, Twiss S (1995) Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Molecular Ecology*, **4**, 653–662.

Amos W, Hoelzel AR (1991) Long-term preservation of whale skin for DNA analysis. *Reports of the International Whaling Commission Special Issue*, **13**, 99–103.

Anderson S, Baker J, Primer J, Baird A (1979) Mortality in Grey seal pups: incidence and causes. *Journal of Zoology, London*, **189**, 407–417.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Baker J (1984) Mortality and morbidity in grey seal pups (*Halichoerus grypus*). Studies on its causes, effects of environment, the nature and sources of infectious agents and the immunological status of pups. *Journal of Zoology, London*, **203**, 23–48.

Baker J (1988) Further studies on grey seal (*Halichoerus grypus*) pup mortality on North Rona. *British Veterinary Journal*, **144**, 497–506.

Baker J, Baker R (1988) Effects of environment on grey seal (*Halichoerus grypus*) pup mortality: studies on the Isle of May. *Journal of Zoology, London*, **216**, 529–537.

Baker C, Vant M, Dalebout M, Lento G, O'Brien S, Yuhki N (2006) Diversity and duplication of *DQB* and *DRB*-like genes of the MHC in baleen whales (suborder: Mysiceti). *Immunogenetics*, **58**, 283–296.

Bean K (2005) Genetic analysis of mate choice, offspring survival and pathogen susceptibility in natural populations. PhD Thesis, University of St Andrews. 195 pp.

Bean K, Amos W, Pomeroy P, Twiss D, Coulson N, Boyd I (2004) Patterns of parental relatedness and pup survival in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, **13**, 2365–2370.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings: Biological Sciences*, **263**, 1619–1626.

Biedrzycka A, Radwan J (2008) Population fragmentation and major histocompatibility complex variation in the spotted suslik, *Spermophilus suslicus*. *Molecular Ecology*, **17**, 4801–4811.

Bowen L, Aldridge B, Gulland F *et al.* (2002) Molecular characterization of expressed *DQA* and *DQB* genes in the California sea lion (*Zalophus californianus*). *Immunogenetics*, **54**, 332–347.

Bowen W, McMillan J, Mohn R (2003) Sustained exponential population growth of grey seals at Sable Island, Nova Scotia. *ICES Journal of Marine Science*, **60**, 1265–1274.

Bowen L, Aldridge BM, Delong R *et al.* (2006) MHC gene configuration in geographically disparate populations of California sea lions (*Zalophus californianus*). *Molecular Ecology*, **15**, 529–533.

Boyd JM, Campbell RN (1971) The grey seal (*Halichoerus grypus*) at North Rona, 1959 to 1968. *Journal of Zoology*, **164**, 469–512.

Brown JH, Jardetzky TS, Gorga JC *et al.* (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.

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

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256**, 50–52.

Ekblom R, Aresæther S, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439–1451.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fraser BA, Neff BD (2010) Parasite mediated homogenizing selection at the MHC in guppies. *Genetica*, **138**, 273–278.

Fraser BA, Ramnarine IW, Neff BD (2010a) Selection at the MHC class IIB locus across guppy (*Poecilia reticulata*) populations. *Heredity*, **104**, 155–167.

Fraser BA, Ramnarine IW, Neff BD (2010b) Temporal variation at the MHC class IIB in wild populations of the guppy (*Poecilia reticulata*). *Evolution*, **64**, 2086–2096.

Gaggiotti OE, Jones F, Lee WM, Amos W, Harwood J, Nichols RA (2002) Patterns of colonization in a metapopulation of grey seals. *Nature*, **416**, 424–427.

Goodman SJ (1997) Dinucleotide repeat polymorphisms at seven anonymous microsatellite loci cloned from the European harbour seal (*Phoca vitulina vitulina*). *Animal Genetics*, **28**, 310–311.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Grimholt U, Larsen S, Nordmo R *et al.* (2003) MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility complex class I and class II loci. *Immunogenetics*, **55**, 210–219.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

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

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hill AVS, Allsopp CEM, Kwiatkowski D *et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.

Hoelzel A, Stephens J, O'Brien S (1999) Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds. *Molecular Biology and Evolution*, **16**, 611–618.

Hoffman JI, Boyd ILB, Amos W (2004) Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*. *Evolution*, **58**, 2087–2099.

Hoffman JI, Hanson N, Forcada J, Trathan PN, Amos W (2010) Getting long in the tooth: a strong positive correlation between canine size and heterozygosity in the Antarctic fur seal *Arctocephalus gazella*. *Journal of Heredity*, **101**, 527–538.

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

Hughes A, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 958–962.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.

Klein J, Sato A (2000a) The HLA system: first of two parts. *New England Journal of Medicine*, **343**, 702–709.

Klein J, Sato A (2000b) The HLA system: second of two parts. *New England Journal of Medicine*, **343**, 782–786.

Klein J, Bontrop R, Dawkins R *et al.* (1990) Nomenclature for the major histocompatibility complex of different species: a proposal. *Immunogenetics*, **31**, 217–219.

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 (*Salmon salar*). *Molecular Ecology*, **10**, 2525–2539.

Lento G, Baker S, David V, Yuhki N, Gales N, O'Brien S (2003) Automated single-strand conformation polymorphism reveals low diversity of a major histocompatibility complex class II gene in the threatened New Zealand sea lion. *Molecular Ecology Notes*, **3**, 346–349.

Loiseau C, Richard M, Garnier S *et al.* (2009) Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, **18**, 1331–1340.

McCallum HI, Kuris A, Harvell CD, Lafferty KD, Smitt GW, Porter J (2004) Does terrestrial epidemiology apply to marine systems? *Trends in Ecology and Evolution*, **19**, 585–591.

McClelland E, Penn D, Potts W (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, **71**, 2079–2086.

Mikkelsen B (2007) Present knowledge of grey seals (*Halichoerus grypus*) in Faroese waters. *NAMMCO Science Publications*, **6**, 79–84.

Miller KM, Kaukinen KH, Beacham TD, Withler RE (2001) Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica*, **111**, 237–257.

Miller KM, Winton JR, Schulze AD, Purcell MK, Ming TJ (2004) Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. *Environmental Biology of Fishes*, **69**, 307–316.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321–3323.

van Oosterhout C, Joyce DA, Cummings SM (2006) Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. *Heredity*, **97**, 111–118.

Penn D, Damjanovich K, Potts W (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11260–11264.

Pomeroy P, Twiss S, Redman P (2000) Philopatry, site fidelity and local kin associations within grey seal breeding colonies. *Ethology*, **106**, 899–919.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rice WR (1989) Analyzing table of statistical tests. *Evolution*, **43**, 223–225.

Sambrook J, Fritsch E, Maniatis T (1989) Extraction with phenol:chloroform. In: *Molecular Cloning: A Laboratory Manual*, 2nd ed. (eds Ford N, Nolan C, Ferguson M), Vol. 3, pp. E.3–E.4. Cold Spring Harbor Laboratory Press, Plainview.

Slade R (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proceedings of the Royal Society, Series B*, **249**, 163–171.

Slade RW, McCallum HI (1992) Overdominance vs. frequency-dependent selection at MHC loci. *Genetics*, **132**, 861–862.

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, **2**, 16–33.

Special Committee on Seals (SCOS) (2007) Scientific Advice on Matters related to the Management of Seal Populations: 2007. Main Advice 2007. 93 pp.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society, Series B*, **277**, 979–988.

Weber D, Stewart B, Schienman J, Lehman N (2004) Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Molecular Ecology*, **13**, 711–718.

Wegner K, Reusch T, Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16**, 224–232.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Westerdahl H, Hansson B, Bensch S, Hasselquist D (2004) Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *Journal of Evolutionary Biology*, **17**, 485–492.

Worthington Wilmer J, Allen PJ, Pomeroy PP, Twiss SD, Amos W (1999) Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, **8**, 1417–1429.

K.C. is a graduate student interested in the role of genetic variation in fitness of marine mammal populations. This work is part of her Masters carried out at Cambridge University. J.I.H. is a postdoctoral research fellow interested in evolutionary and conservation genetics, primarily in pinnipeds. L.A.K. is a Cambridge Reader in Primate Genetics and Behaviour interested in primate immunogenetics and molecular ecology. J.H. is a Professor at the Scottish Oceans Institute interested in population dynamics, genetics, and epidemiology of vertebrates, primarily marine mammals. W.A. is a Cambridge Professor of Evolutionary Genetics interested in the effects of genetic variation on survival and reproductive success.

Data accessibility:

DNA sequences: GenBank accessions HQ456122–HQ456126.

Data deposited at Dryad: doi:10.5061/dryad.8175.