Evidence for Balancing Selection at the Major Histocompatibility Complex in a Free-Living Ruminant

S. Paterson

Evidence for selective maintenance of genetic diversity at the major histocompatibility complex (MHC) was investigated in an unmanaged population of Soay sheep on the island of Hirta, St. Kilda, Scotland. Animals were sampled as newborn lambs and between 887 and 1209 individuals were typed at each of five microsatellite markers located either within or flanking the ovine MHC. Markers located within the MHC showed high levels of linkage disequilibrium with each other but not with flanking markers. Hardy-Weinberg proportions were found for all loci; however, two of the three markers within the MHC showed relatively even allele frequency distributions that were unlikely to have resulted from neutrality and suggest the action of recent balancing selection. Sequence polymorphism was examined within DRB, a class II gene immediately adjacent to one of the microsatellite markers, and high correlation of sequence polymorphism with microsatellite length variation was found. An excess of nonsynonymous substitution compared to synonymous substitution was found, indicating the action of balancing selection favoring novel MHC variants and hence increased diversity over a longer time period.

Molecular techniques have steadily increased the known level of genetic variation present in natural populations, but the processes that can act to maintain genetic diversity are still poorly understood. For example, balancing selection has been relatively poorly documented in nature, the few known examples including the sickle-cell hemoglobin polymorphism in humans (Allison 1964), color and band polymorphism in *Cepea nemoralis* (Jones et al. 1977), warfarin resistance in rats (Bishop 1981), and alcohol dehydrogenase (*Adh*) polymorphism in *Drosophila melanogaster* (van Delden 1982). The systems outlined above, however, are confined to particular species, and in some cases little is known about their molecular basis (although it is noted that sickle cell anemia in humans and *Adh* in *D. melanogaster* are exceptionally well characterized at the molecular level). In order to study the maintenance of genetic diversity in the natural environment, a system is required which is common to a large group of organisms and which is well understood at the molecular and physiological level. In this respect, the major histocompatibility complex (MHC) of vertebrates has become an evolutionary genetic paradigm (Hedrick 1994).

The MHC of vertebrates comprises a group of closely linked genes which play a central role in immune surveillance and response (Klein 1986; Weir and Stewart 1993). MHC molecules loaded with peptides derived from invading pathogens are presented on the surface of infected cells or of specialized cells of the immune system. These unique MHC-peptide combinations are recognized by the immune system to produce a highly effective and specific response against foreign pathogens (Benacerraf 1981). High levels of diversity and polymorphism at certain loci within the complex have been documented in a wide range of vertebrate species (Klein 1986).

A number of indirect lines of evidence suggest a selective force maintaining MHC diversity. First, where genealogies of MHC allele sequences have been constructed from a number of species [for example, primates (Klein et al. 1993a) and felines (Yuhki and O'Brien 1997)] it is observed that the divergence of allelic MHC lineages predates the speciation event giving rise to separate taxa. This suggests the action of some form of balancing selection over long periods of evolutionary time. Second, comparisons of allelic sequences present within mice and human populations indicate that the rate of nonsynonymous substitution exceeds the rate of synonymous

From the Department of Genetics, University of Cambridge, Downing St., Cambridge CB2 3EH, England. I thank the numerous staff and volunteers who have helped collect samples and data on St. Kilda, especially A. Robertson, A. MacColl, and J. Pilkington; Scottish Natural Heritage and the Natural Trust for Scotland for permission to work on St. Kilda and the Royal Artillery Range (Hebridies) and the Royal Corps of Transport for logistical assistance; Bill Amos, Simon Goodman, Katrina Lythgoe, and Josephine Pemberton for comments during the preparation of this manuscript. This research was funded by BBSRC and NERC. Sequences described in this article have been submitted to the EMBL database, accession nos. Y10245–Y10249. Address correspondence to the author at ICAPB, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, United Kingdom, or e-mail: s.paterson@ed.ac.uk.

 89:289–294

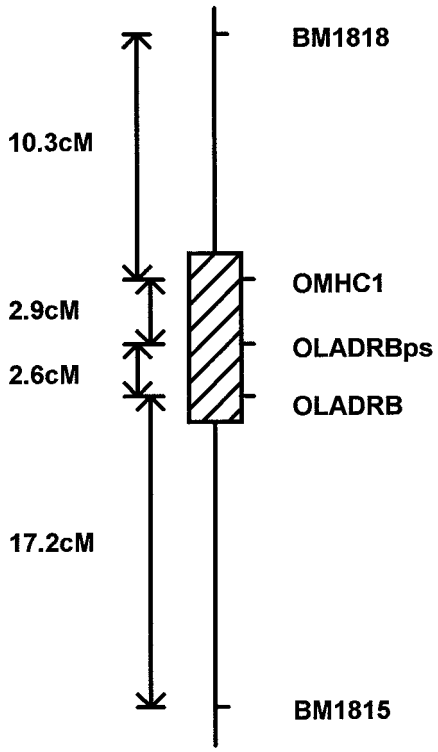

Figure 1. Schematic (not to scale) of part of sheep chromosome 20 showing MHC region (box) and microsatellite markers used in this study. Genetic distances and nomenclature are taken from Crawford et al. (1995).

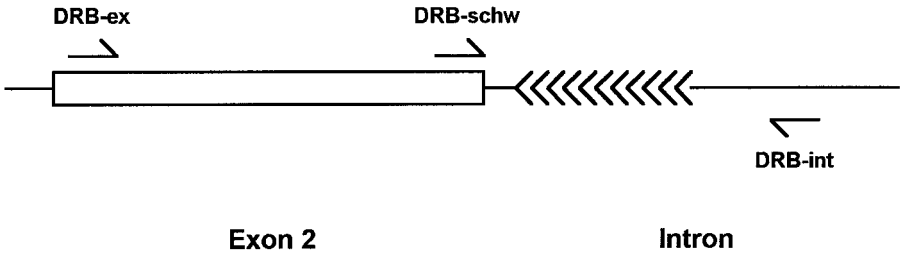

Figure 2. Schematic of DRB gene showing exon 2 and adjacent 3′ intron containing the OLADRB microsatellite (– – –<<<– – –). The primer pair DRB-ex and DRB-int amplify an approximately 480–550 bp fragment corresponding to Ovar-DRB exon 2 plus adjacent OLADRB microsatellite. The primer pair DRB-schw and DRB-int amplify an approximately 200–270 bp fragment corresponding to the OLADRB microsatellite.

substitution at the antigen presenting site (APS), thus favoring new MHC variants and increasing diversity (Hughes et al. 1994; Hughes and Nei 1988, 1989). Finally, in human populations the large numbers of alleles present at MHC loci show a relatively even distribution, leading to higher levels of heterozygosity than may be explained under neutral theory (Hedrick and Thomson 1983) and suggesting the action of recent balancing selection. So far, however, there is a paucity of population-based studies of MHC diversity, with these last two lines of evidence being confined to a relatively small number of taxa, and often involving species whose recent genetic history has been heavily influenced by humans.

In this article I examine patterns of diversity found at the MHC in the unmanaged population of Soay sheep living on St. Kilda. I find evidence of balancing selection from allele frequency distributions and expressed sequence polymorphism.

Materials and Methods

Study Site, Animals, and Sampling

The archipelago of St. Kilda lies 45 miles west of the Outer Hebrides, Scotland (57°49′N, 08°34′W), consisting of the islands of Hirta, Soay, Dun, and Boreray. Soay sheep are a primitive breed of sheep, resembling the neolithic domestic sheep first brought to Britain about 5000 B.C. They may have been introduced to St. Kilda as early as 2000 B.C. and since historic times have been restricted to the uninhabited island of Soay (Jewell et al. 1974).

In 1932, following the evacuation of the human population 2 years previously, 107 Soay sheep (20 rams, 44 ewes, 21 ewe lambs, and 22 castrated ram lambs) were introduced from Soay (99 ha) to the larger island of Hirta (638 ha; Jewell et al. 1974). Numbers increased rapidly and since 1952 have fluctuated between 600 and 1800 animals. The population of sheep within the study area (approximately 200 ha) constitutes about one-third of the total Hirta population, fluctuating in size between 200 and 550 animals and correlating closely with the rest of the island population (Clutton-Brock et al. 1991, 1992). Since 1985 at least 95% of lambs born within the study area have been caught and individually tagged. Blood samples and ear punches for genetic analysis were taken from lambs at this time.

Microsatellite Analysis

DNA was isolated from ear punches or white blood cells by standard proteinase K digestion, followed by phenol/chloroform extraction and ethanol precipitation (Sambrook et al. 1989). After preliminary trials of a total of 13 published ovine or bovine microsatellites mapped to the MHC region (Bishop et al. 1994; Crawford et al. 1995), five loci were selected for this study on the basis of their map locations and their variability in the study population. The location of these markers relative to each other is shown in Figure 1, with genetic distances based on a mapping flock of domestic sheep described in Crawford et al. (1995). The markers OLADRB (Schwaiger et al. 1993a) and OLADRBps (Blattman and Beh 1992) are located within MHC class II DRB expressed and nonexpressed genes, respectively, OMHC1 is located within the MHC class I region (Groth and Wetherall 1994) and BM1815 and BM1818 (Bishop et al. 1994) were used as flanking marker controls. Animals were sampled as newborn lambs between 1988 and 1994 and between 887 and 1209 individuals were typed at each of these five microsatellite markers. Sample sizes varied between loci due to time constraints on this project and due to the ease with which some loci amplified relative to others.

Microsatellite primer sequences are described elsewhere (Crawford et al. 1995), with the exception of OLADRB for which I used specific primers designed to allow strong amplification from sheep sequence (DRB-schw, 5′-TG$^{T}/_{G}$GCAGCGGCGAGGTGAG and DRB-int, 5′-CGTACCCAGA$^{T}/_{G}$TGAGTGAAGTATC; see Figure 2). PCR was carried out in a Hybaid Omnigene thermal cycler in 10 μl reactions overlaid with mineral oil containing: 1× PARR buffer (Cambio, UK); 10% DMSO; 2.0 mM $MgCl_2$; 0.1 mM dGTP, dATP, and dTTP; 0.01 mM dCTP; 0.5 μCi α-^{32}P dCTP; 0.25 U *Taq* polymerase; 4 pmol of each primer; 50 ng genomic DNA. Standard cycling conditions were 93°C for 2 min followed by 7 cycles of 93°C for 30 s, 52°C for 60 s, 72°C for 30 s, and 25 cycles of 93°C for 30 s, 54°C for 60 s, 72°C for 30 s with the following modifications of annealing temperature: OLADRB, 46°C and 54°C; BM1815, 50°C and 56°C; BM1818, 48°C and 50°C. PCR products were separated on 6% denaturing polyacrylamide gels and visualized by exposure to Kodak X-OMAT™ AR X-ray film for 12–144 h depending on signal strength (Sambrook et al. 1989).

Cloning and Sequencing Ovar-DRB Exon 2

The gene DRB encodes a class II MHC product involved in antigen presentation for which high levels of diversity have

Table 1. Population data for the five microsatellite loci screened in the Soay population

Marker	No. of animals screened	No. of alleles	Heterozygosity	Hardy-Weinberg exact text (*P* value)
MB1815	961	3	0.51	.17
OLADRB	1,209	8	0.80	.16
OLADRBps	887	6	0.77	.97
OMHC1	1,025	5	0.58	.27
BM1818	893	6	0.66	.38

For each locus, the number of animals typed, the number of alleles found, heterozygosity and Hardy-Weinberg exact *P* values are shown. The loci OLADRB, OLADRBps, and OMHC1 (shown in boldface) lie within the ovine MHC, BM1815 and BM1818 are flanking loci (see Figure 1).

been found in humans, mice, cattle, and sheep (Klein et al. 1993b; Schwaiger et al. 1993b). The vast majority of the polymorphism found in this gene is located in exon 2 which encodes the antigen presentation site (Brown et al. 1988; Klein et al. 1993b). The microsatellite OLADRB used in this study (see above) is located within an intron approximately 30 bp 3′ to exon 2 (Figure 2; Schwaiger et al. 1993a).

To assay sequence diversity within this region, DRB exon 2 was PCR amplified, purified, and cloned from 15 animals (13 homozygotes and 2 heterozygotes) representing between four and six copies of each of the six common microsatellite length variants at OLADRB. DRB exon 2 and the flanking microsatellite sequence were amplified from genomic sheep DNA in 50 μl reactions in 0.5 ml Eppendorf tubes containing the following (Schwaiger et al. 1994): 150 ng genomic DNA; 2.0 mM $MgCl_2$; 20 pmol primer DRB-ex (TCTCTGCAGCACATTTCCTGG), 20 pmol primer DRB-int (CGTACCCAGA-$^{G}/_{T}$GAGTGAAGTATC), PARR buffer (Cambio), 2 U *Taq* polymerase, 10% DMSO, and 100 μM dNTPs. Reactions were overlaid with light mineral oil and temperature cycled in a Hybaid Omnigene as follows: 93°C for 2 min; 30 cycles of 46°C for 50 s, 72°C for 50 s, 93°C for 50 s; 46°C for 50 s; 72°C for 5 min. PCR products were then gel purified on 1.5% agarose gels using Promega Wizard preps following manufacturer's instructions. Purified PCR product was then ligated into Promega T-vector in a 10 μl reaction containing 10–30 ng insert DNA; 30 mM Tris-HCl (pH 7.5);10 mM $MgCl_2$; 10 mM DTT; 10 mM ATP; 50 ng T-vector; and 1 U T4 ligase. Reactions were incubated at 15°C for 4 h followed by 4°C overnight. Ligations were then transformed into *E. coli* JM109 cells by electroporation. Cells were plated on Amp/IPTG L-agar plates. Positive clones were screened to ensure that PCR products of the desired length were inserted and to determine orientation of insert. This was done by restriction digest with *Pvu*II, PCR amplification of OLADRB microsatellite and PCR amplification of DRB exon 2 plus microsatellite. At least 6 clones for OLADRB homozygotes and at least 10 clones for OLADRB heterozygotes (5 clones for each allele) were sequenced using dye-labeled M13 (-21) forward primer and loaded on an ABI 377 automatic sequencer.

Statistical Treatment

Disequilibria. Hardy-Weinberg assumptions for each locus and linkage equilibrium assumptions for pairs of loci were tested by the Markov-chain method in the population genetic package Genepop 1.2 (Raymond and Rousset 1995), which generates exact tests for multiple alleles (Guo and Thompson 1992). Disequilibrium analysis was conducted within cohorts in case of any possible biases due to temporal fluctuations in allele frequencies.

The relative importance of linkage and genetic drift between cohorts in contributing to disequilibrium were tested by calculating variance components of linkage disequilibrium following the method of Ohta (1982). D^2_{IS} and D^2_{ST} denote the within and between cohort components of disequilibrium and D^2_{IT} denotes the total disequilibrium:

$$D^2_{IS} = E\left\{\sum_{i,j} (g_{ij,k} - x_{i,k}y_{j,k})^2\right\}, \quad (1)$$

$$D^2_{ST} = E\left\{\sum_{i,j} (x_{i,k}y_{j,k} - \bar{x}_i\bar{y}_j)^2\right\}, \text{ and} \quad (2)$$

$$D^2_{IT} = E\left\{\sum_{i,j} (g_{ij,k} - \bar{x}_i\bar{y}_j)^2\right\}, \quad (3)$$

where $g_{ij,k}$ is the frequency of chromosomes in cohort k carrying alleles i and j at loci A and B, respectively, and $x_{i,k}$ and $y_{j,k}$ are the frequencies of A_i and B_j in cohort k, respectively. Population means are given by $\bar{g}_{ij} = 1/n\ \Sigma_{k=1}^{n}\ g_{ij,k}$, $\bar{x}_j = 1/n\ \Sigma_{k=1}^{n}\ x_{i,k}$, and $\bar{y}_j = 1/n\ \Sigma_{k=1}^{n}\ y_{j,k}$, where n is the number of cohorts.

These components (D^2_{IS}, D^2_{ST}, and D^2_{IT}) were subsequently adjusted for heterozygosity such that

$$D^* = \frac{D^2}{H_A H_B}, \quad (4)$$

where $H_A = \Sigma_i\ \bar{x}_i^2$ and $H_B = \Sigma_i\ \bar{y}_i^2$ (Hedrick 1994).

Allele frequency distributions. Balancing selection was tested for by using Watterson's (1978) homozygosity test ($F = \Sigma_{i=1}^{n} p_i^2$) for sampling n alleles at given allele frequencies under a neutral model and relating the calculated F value for each locus to published significance levels (Watterson 1978).

Synonymous versus nonsynonymous substitutions. Nucleotides believed to be involved in antigen presentation were identified by comparison with the model produced by Brown et al. (1988, 1993a,b). Genetic distances for synonymous (d_S) and non-synonymous substitutions (d_N) were calculated using the Jukes-Cantor correction for multiple hits (Kumar et al. 1993). The difference between these distances ($d = \bar{d}_N - \bar{d}_S$) was tested against the t distribution with infinite degrees of freedom according to the test statistic

$$t = d/s(d), \quad (5)$$

where $s(d)$ is the standard error of d, given by $s(d) = [\text{var}(\bar{d}_S) + \text{var}(\bar{d}_N)]^{1/2}$ (Kumar et al. 1993).

Results

Microsatellite Variation

Summary data for the five microsatellite loci are shown in Table 1. All loci were found to be in Hardy-Weinberg proportions.

Linkage Disequilibrium

Microsatellites are believed to be selectively neutral (Bruford and Wayne 1993) and are used in this study to act as markers for any adjacent sites within the MHC under selection. To detect any effects of balancing selection, such an approach requires linkage disequilibrium between neutral markers and selected sites within the MHC. Patterns of linkage disequilibrium were analyzed to justify the use of such an approach. No attempt is made to conclude whether selection is an important factor in the generation of observed patterns of linkage disequilibrium (see Discussion).

Table 2. Variance components of linkage disequilibrium for all pairwise comparisons partitioned into within and between cohort components

Loci compared	Distance between loci (cM)	Within cohort component D^2_{IS} (D^*_{IS})	Between cohort component D^2_{ST} (D^*_{ST})	Total population D^2_{IT} (D^*_{IT})
OLADRB-OLDRBps	2.6	0.11540 (0.1792)	0.00257 (0.0040)	0.12482 (0.1938)
OLADRBps-OMHC1	2.9	0.05780 (0.1214)	0.00392 (0.0082)	0.06839 (0.1436)
OLADRB-OMHC1	5.5	0.05294 (0.1087)	0.00280 (0.0057)	0.06610 (0.1357)
OMHC1-BM1818	10.3	0.00609 (0.0147)	0.00264 (0.0064)	0.01627 (0.0393)
OLADRBps-BM1818	13.2	0.01159 (0.0212)	0.00263 (0.0048)	0.02060 (0.0377)
OLADRB-BM1818	15.8	0.01115 (0.0199)	0.00218 (0.0039)	0.02044 (0.0365)
OLADRB-BM1815	17.2	0.00690 (0.0158)	0.00344 (0.0079)	0.01914 (0.0438)
OLADRBps-BM1815	19.8	0.00822 (0.0193)	0.00463 (0.0109)	0.01930 (0.0453)
OMHC1-BM1815	22.7	0.01267 (0.0392)	0.00323 (0.0100)	0.02616 (0.0809)
BM1815-BM1818	33.0	0.00407 (0.0110)	0.00346 (0.0094)	0.01692 (0.0457)

Genetic distances between loci shown. The figures adjusted for heterozygosity are shown in brackets. Pairwise comparisons between loci within the MHC are shown in boldface.

Significant disequilibrium (within cohorts) was found between pairs of markers within the MHC ($P < .001$ in all cases) but not for comparisons involving flanking markers. As shown by the variance components presented in Table 2, the degree of total disequilibrium (D^2_{IT}) between markers within the MHC was three to six times higher than between MHC and flanking markers, with the highest disequilibrium observed between the two class II markers OLADRB and OLADRBps. Disequilibrium within cohorts (D^2_{IS}) was up to 10 times higher between markers in the MHC compared with disequilibrium between MHC and flanking markers. Compared to total disequilibrium (D^2_{IT}), disequilibrium within cohorts (D^2_{IS}) more accurately reflects disequilibrium due to linkage, this measure (D^2_{IS}) is independent of disequilibrium due to temporal variation in allele frequencies (D^2_{ST}). The effect of genetic drift in allele frequencies between cohorts upon disequilibrium (D^2_{ST}) was approximately equal for all locus pairs.

Allele Frequency Distributions

Individual allele frequencies and the results of Watterson's (1978) test for homozygosity are shown in Table 3. Two markers within the class II region of the MHC, OLADRB and OLADRBps, showed very even allele frequency distributions ($P < .01$ in both cases). These are unlikely under neutrality and suggest the action of balancing selection maintaining heterozygosity in this region of the MHC. Allele frequency distributions at the class I marker OMHC1 and the two flanking markers showed no deviations from neutral expectations.

Expressed DRB Exon 2 Sequence Variation

Five sequence variants were observed in the animals sampled (Figure 3), although I note that a larger sample screened by SSCP or heteroduplex analysis may have detected other, rarer variants. All the observed sequence variants could be distinguished on the basis of adjacent microsatellite length variation. The sequence variants were designated OAMHC205, OAMHC213, OAMHC257, OAMHC263, and OAMHC276 (EMBL accession numbers Y10245–Y10249) due to their association with OLADRB microsatellite length variants 205, 213, 257, 263, and 276, respectively. One sequence variant, OAMHC263, was associated with both the 263 and 267 microsatellite length variants at OLADRB. Comparison with all submitted sheep DRB sequences on the EMBL database indicated that two of these sequences, OAMHC257 and OAMHC276, appear identical to alleles sequenced previously from domestic flocks (EMBL accession numbers U00206 and Z11520, respectively).

Codons involved in antigen presentation are indicated in Figure 3. Within these sites, the average synonymous substitution rate ($\bar{d}_S$) was found to be 0.0939 ± 0.0705 SE and the average nonsynonymous substitution rate ($\bar{d}_N$) was found to be 0.2945 ± 0.0671 SE. $t = d/s(d)$ was found to equal 2.06 (significant at the 5% level), indicating that the rate of nonsynonymous substitution exceeds the rate of synonymous substitution, suggesting the action of selection upon the class II MHC antigen presenting site of Soay sheep (Hughes and Nei 1988, 1989).

Discussion

I used three microsatellite loci located within the MHC and two flanking marker controls to rapidly assay genetic variation in the MHC region of Soay sheep. All loci were in Hardy-Weinberg proportions. The three markers located within the MHC showed high levels of linkage disequilibrium, but disequilibrium was not found between MHC and flanking markers.

The demographic history of the Soay

Table 3. Allele sizes and frequencies for microsatellite markers used in this study

	BM1815		**OLADRB**		**OLADRBps**		**OMHC1**		BM1818	
Marker	Size (bp)	Frequency	Size (bp)	Frequency	Size (bp)	Frequency	Size (bp)	Frequency	Size (bp)	Frequency
Allele 1	141	0.041	205	0.211	265	0.108	184	0.111	258	0.085
Allele 2	146	0.479	213	0.113	271	0.124	192	0.096	260	0.331
Allele 3	153	0.480	257	0.236	273	0.258	198	0.598	264	0.064
Allele 4	—	—	263	0.133	277	0.086	202	0.079	276	0.005
Allele 5	—	—	267	0.158	281	0.304	206	0.116	278	0.086
Allele 6	—	—	269	0.001	283	0.121	—	—	280	0.001
Allele 7	—	—	276	0.141	—	—	—	—	282	0.428
Allele 8	—	—	287	0.008	—	—	—	—	—	—
Observed proportion homozygotes	0.494		0.204		0.212		0.419		0.345	
Watterson's test statistic (Σp_i^2)	0.462 ($k = 3$)		0.187 ($k = 8$)		0.208 ($k = 6$)		0.399 ($k = 5$)		0.312 ($k = 7$)	
Probability	n.s.		$p < .01$		$p < .01$		n.s.		n.s.	

MHC markers shown in boldface. The proportion of homozygotes, Watterson's (1978) test statistic for balancing selection (Σp_i^2), and the probability of the observed frequency distribution under neutrality are shown for each marker. Effects of balancing selection are observed at OLADRB and OLADRBps, despite Hardy-Weinberg proportions for all loci. n.s. = not significant.

a)

```
allele                                           30                                      60
OAMHC205    TAT CGT AAG AGC GAG TGT CGT TTC TCC AAC GGG ACG GAG CGG GTG CGG TTC CTG GAC AGA
OAMHC213    ... A.. ... ... ... ... .A. ... .T. ... ... ..C ... ... ... ... ... ... ... ...
OAMHC257    ... AC. ... .AA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OAMHC263    ... AC. ... .AA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OAMHC276    ... AC. ... .AA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

allele                                           90                                     120
OAMHC205    TAC TTC TAT AAT GGA GAA GAG AAC GTG CGC TTC GAC AGC GAC TGG GGC GAG TAC CGA GCG
OAMHC213    ... .A. AC. ..C ... ... ... ... ... ... ... ... ... ... ... ... ... .T. ..G ...
OAMHC257    ... ... C.. ... ... ... ... .C. C.. ... ... ... ... ... ... ... ... ... ... ...
OAMHC263    ... ... C.. ... ... ... ... .C. C.. ... ... ... ... ... ... ... ... ... ... ...
OAMHC276    ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ...

allele                                          150                                     180
OAMHC205    GTG GCC GAG CTG GGG CGG CCG GAC GCC AAG TAC TGG AAC AGC CAG AAG GAC TTC CTG GAG
OAMHC213    ... A.. ... ... ... ... ... ... ..T G.. C.A ... ... ... ... ... ... ... ... ...
OAMHC257    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OAMHC263    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OAMHC276    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G A.. ... ...

allele                                          210                                     240
OAMHC205    CGG AAG CGG GCC AAT GTG GAC ACG TAC TGC AGA CAC AAC TAC GGG GTC GGT GAG AGT TTC
OAMHC213    A.C .G. A.. A.. GCG ... ... ... ... ... ... ... ... ... ... ... TT. ... ... ...
OAMHC257    ... GC. ... ... GCC ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
OAMHC263    ... GC. ... ... GCC ... ... ... ... ... ... ... ... ... ... ... AT. ... ... ...
OAMHC276    ... .G. ... A.. G.G ... ... ... ... ... ... ... ... ... ... ... AT. ... ... ...

allele                            ][ Intron->   270
OAMHC205    ACT GTG CAG CGG CGA GGT GAG CGC GGG GGT GGG CGG CCA ATG TGG AGC
OAMHC213    .G. ... ... ... ... ... ... ... ... ... ... .A. ... G.. ... ...
OAMHC257    .G. ... ... ... ... ... ... ... ... ... ... ... ... G.. ... ...
OAMHC263    ... ... ... ... ... ... ... ... ... ... ... ... ... C.. ... ...
OAMHC276    .G. ... ... ... ... ... ... ... ... ... ... ... ... G.. ... ...
```

b)

```
allele              10         20         30         40         50         60         70
OAMHC205    YRKSECRFSN GTERVRFLDR YFYNGEENVR FDSDWGEYRA VAELGRPDAK YWNSQKDFLE RKRANVDTYC
OAMHC213    .S....H.F. .......... .YT....... .......F.. .T.......E Q......... SR.TA.....
OAMHC257    .T.K...... .......... ..H....TL. .......... .......... .......... .A..A.....
OAMHC263    .T.K...... .......... ..H....TL. .......... .......... .......... .A..A.....
OAMHC276    .T.K...... .......... .......Y.. .......... .......... ......EI.. .R.TE.....

allele              80
OAMHC205    RHNYGVGESF TVQRR
OAMHC213    ......F... S....
OAMHC257    .......... S....
OAMHC263    ......I... .....
OAMHC276    ......I... S....
```

Figure 3. Alignment of **(a)** nucleic acid and **(b)** predicted protein sequence for the five sequence variants found at DRB exon 2 in the Soay population. Sequences are arbitrarily compared against OAMHC205. Codons/residues believed to be involved in antigen presentation are highlighted in bold. The boundary between exonic and intronic sequence is indicated at nucleic acid position 258/259.

sheep on St. Kilda is complex, with an initial founder event (see Materials and Methods), overlapping generations, fluctuating population size and variation in reproductive success (Clutton-Brock et al. 1991, 1992; Pemberton et al. 1996). It is unclear the extent to which selection may have contributed to the level of disequilibrium observed within the population since the influence of these demographic processes upon linkage disequilibrium are difficult to predict. However, regardless as to cause, the linkage disequilibrium found within the MHC (but not between MHC and flanking markers) justifies my approach of using neutral microsatellite markers, since any sites within the MHC under the influence of balancing selection is likely to be detected using the markers in this study. Moreover, effects of balancing selection should not be detected by flanking marker controls.

I investigated the role of balancing selection within the MHC region of Soay sheep by comparing allele frequency distributions against neutral expectations (Watterson 1978). Of the five loci screened in the Soay sheep population, allele frequency distributions of the two markers within the MHC class II region, OLADRB and OLADRBps, showed highly significant deviations from neutral expectations ($P <$ .01 in both cases). By contrast, the allele frequency distribution of the class I marker OMHC1 and the two flanking control loci showed no significant deviations. The recent founder event in the Soay sheep population could act to eliminate rare alleles and lead to even allele frequencies independent of selection. This, however, is an unlikely explanation for the results observed in this study, since unusually even allele frequencies are observed only at MHC loci, with neither flanking markers at the MHC nor microsatellites distributed elsewhere in the Soay genome (Pemberton et al., in preparation) showing markedly even allele frequency distributions. The most likely explanation for the pattern of results observed in this study is balancing selection at sites within the MHC acting over the recent history of the Soay sheep population.

Failure to detect balancing selection at OMHC1 may suggest that an important selected site lies within the MHC class II region, but that sufficient recombination between the class I and II regions exists to disrupt the effects of balancing selection in maintaining genetic variation over the entire MHC region. Indeed, the analysis of linkage disequilibrium variance components implies that linkage disequilibrium within the class II region is around twofold higher than that found between class I and II regions. There is, however, insufficient data to support this conclusion. Particular alleles at OMHC1 may be associated with more than one sequence variant which would reduce my ability to detect balancing selection within the class I region of the MHC using the microsatellite marker OMHC1. Development of further microsatellite markers or direct genotyping of functional polymorphism within the MHC would help increase the resolution of this study.

I also investigated sequence variation in DRB, an expressed gene product involved in antigen binding. As in other studies of this region in ruminants (Ammer et al. 1992; Schwaiger et al. 1993a,b), high correlation was found between sequence polymorphism and adjacent microsatellite length variation. There was no loss of information by typing microsatellite variation relative to sequence analysis, indicating that microsatellite analysis represents a viable approach to assay genetic variation at the MHC. Indeed, one sequence variant OAMHC263 was associated with two microsatellite length variants at OLADRB, and it is possible that microsatellite variation is marking functional polymorphism outside DRB exon 2 sequence.

In common with previous mouse and human studies (Hughes and Nei 1988, 1989), I detected evidence of balancing selection in patterns of substitution within the antigen presentation site. The rate of evolution appears much higher for nonsynonymous than for synonymous substitutions, indicating selection for increased diversity at residues associated with antigen presentation to the immune system. This result suggests that selection has operated to shape extant MHC variation in the population over a far longer time scale than is indicated by the allele frequency results. These results can be compared to recent studies of domestic cattle (Mikko and Andersson 1995a), American bison (Mikko et al., in press), and moose (Mikko and Andersson 1995b). In common with this study, these populations show evidence for selection over evolutionary time

in the pattern of nucleotide substitution in extant DRB sequences despite either domestication in cattle, which might act to preclude selection in their recent history, or bottleneck events in bison and moose, which would tend to reduce extant MHC variation.

This study represents one of the most extensive investigations into MHC diversity within a free-living, nonhuman population to date. In common with studies on human and mouse populations, indirect evidence suggests that diversity at the MHC in Soay sheep is nonneutral and is likely to be maintained by balancing selection. The mechanism behind this balancing selection is as yet unclear. Both heterozygote advantage (Doherty and Zinkernagel 1975) and frequency-dependent selection in response to parasites (Hamilton 1980; Hill et al. 1991) and reproductive decisions reducing inbreeding (Potts et al. 1991, 1994) have been proposed as mechanisms able to maintain genetic diversity in vertebrates at the MHC and across the genome. Further work in populations such as the Soay sheep on St. Kilda will be required to elucidate the exact nature of the selective forces acting on the MHC.

References

Allison AC, 1964 Polymorphism and natural selection in human populations. Cold Spring Harbor Symp Quant Biol 29:139–149.

Ammer H, Schwaiger FW, Kammerbauer C, Gomolka M, Arriens A, Lazary S, and Epplen JT, 1992. Exonic polymorphism versus intronic simple repeat hypervariability in MHC-DRB-genes. Immunogenetics 32:330–337.

Benacerraf B, 1981 Role of MHC gene products in immune regulation. Science 212:1229–1238.

Bishop JA, 1981 A neo-Darwinian approach to resistance: examples from mammals. In: Genetic consequences of man-made change (Bishop JA and Cook LM, eds).New York: Academic Press.

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo JY, and Beattie CW, 1994. A genetic-linkage map for cattle. Genetics 136:619–639.

Blattman AN and Beh KJ, 1992. Dinucleotide repeat polymorphism within the ovine major histocompatibility complex. Anim Genet 23:392.

Brown JH, Jardetzky T, Gorga JC, Stern LJ, Strominger JL, and Wiley DC, 1993a. Crystal-structure of the class-II histocompatibility antigen HLA-DR1. J Cell Biochem 517:66–66.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, and Wiley DC, 1993b. 3-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364:33–39.

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, and Wiley DC, 1988. A hypothetical model of the foreign antigen site of class II histocompatibility molecules. Nature 332:845–850.

Bruford MW and Wayne RK, 1993. Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939–943.

Clutton-Brock TH, Price OF, Albon SD, and Jewell PA, 1991. Persistent instability and population regulation in Soay sheep. J Anim Ecol 60:593–608.

Clutton-Brock TH, Price OF, Albon SD, and Jewell PA, 1992. Early development and population fluctuations in Soay sheep. J Anim Ecol 61:381–396.

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG, Beattie AE, Davies K, Maddox JF, Kappes SW, Stone RT, Nguyen TC, Penty JM, Lord EA, Broom JE, Buitkamp J, Schwaiger W, Epplen JT, Matthew P, Matthews ME, Hulme DJ, Beh KJ, McGraw RA, and Beattie CW, 1995. An autosomal genetic-linkage map of the sheep genome. Genetics 140:703–724.

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

Groth DM and Wetherall JD, 1994. Dinucleotide repeat polymorphism within the ovine major histocompatibility complex class I region. Anim Genet 25:61.

Guo SW and Thompson EA, 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.

Hamilton WD, 1980. Sex versus non-sex versus parasite. Oikos 35:282–290.

Hedrick PW, 1994. Evolutionary genetics of the major histocompatibility complex. Am Nat 143:945–964.

Hedrick PW and Thomson G, 1983. Evidence for balancing selection at HLA. Genetics 104:449–456.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, and Greenwood BM, 1991. Common West African HLA antigens are associated with protection from severe malaria. Nature 352:595–600.

Hughes AL, Hughes MK, Howell CY, and Nei M, 1994. Natural selection at the class II major histocompatibility complex loci of mammals. Phil Trans R Soc Lond B 345:359–367.

Hughes AL and Nei M, 1988. Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. Nature 335:167–170.

Hughes AL and Nei M, 1989. Nucleotide substitution at major histocompatibility complex class-II loci—evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962.

Jewell PA, Milner C, and Morton Boyd J, 1974. Island survivors. London: Athlone Press.

Jones JS, Leith BH, and Rawlings P, 1977. Polymorphism in *Cepea*: a problem with too many solutions. Annu Rev Ecol Systemat 8:109–143.

Klein J, 1986. Natural history of the major histocompatibility complex. New York: John Wiley.

Klein J, Satta Y, Ohuigin C, and Takahata N, 1993a. The molecular descent of the major histocompatibility complex. Annu Rev Immunol 11:269–295.

Klein J, Takahata N, and Ayala FJ, 1993b. MHC polymorphism and human origins. Sci Am 269:78–83.

Kumar S, Tamura K, and Nei M, 1993. MEGA: molecular evolutionary genetic analysis. University Park, Pennsylvania: Pennsylvania State University; 11–12.

Mikko S and Andersson L, 1995a. Extensive MHC class II diversity in African and European cattle. Immunogenetics 42:408–413

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

Mikko S, Spencer M, Morris B, Stabile S, Basu T, Stormont C, and Andersson L, 1997. A comparative analysis of MHC DRB3 polymorphism in the American bison (*Bison bison*). J Hered 88:499–503.

Ohta T, 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc Natl Acad Sci USA 79:1940–1944.

Pemberton JM, Smith JA, Coulson TN, Marshall TC, Slate J, Paterson S, Albon SD, and Clutton-Brock TH, 1996. The maintenance of genetic polymorphism in small island populations: large mammals in the Hebrides. Phil Trans R Soc Lond B 351:745–752.

Potts WK, Manning CJ, and Wakeland EK, 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. Nature 352:619–621.

Potts WK, Manning CJ, and Wakeland EK, 1994. The role of infectious-disease, inbreeding and mating preferences in maintaining MHC genetic diversity—an experimental test. Phil Trans R Soc Lond B 346:369–378.

Raymond M and Rousset F, 1995. Genepop (version 1.2)—population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Sambrook J, Fritsch EF, and Maniatis T, 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Schwaiger FW, Buitkamp J, Weyers E, and Epplen JT, 1993a. Typing of artiodactyl MHC genes with the help of intronic simple repeated sequences. Mol Ecol 2:55–59.

Schwaiger FW, Weyers E, Buitkamp J, Ede AJ, Crawford A, and Epplen JT, 1994. Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. Mol Biol Evol 11: 239–249.

Schwaiger FW, Weyers E, Epplen C, Brun J, Ruff G, Crawford A, and Epplen JT, 1993b. The paradox of MHC-DRB exon-intron evolution—alpha-helix and beta-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with beta-sheet codons. J Mol Evol 37:260–272.

van Delden W, 1982. The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: selection at an enzyme locus. Evol Biol 15:187–222.

Watterson GA, 1978. The homozygosity test of neutrality. Genetics 88:405–417.

Weir DW and Stewart J, 1993. Immunology. Edinburgh: Churchill Livingston.

Yuhki N and O'Brien SJ, 1997. Nature and origin of polymorphism in feline MHC class II DRA and DRB genes. J Immunol 158:2822–2833.

Received June 17, 1997
Accepted December 13, 1997

Corresponding Editor: Leif Andersson