Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation?

GUSTAVO A. GUTIERREZ-ESPELETA†‡, PHILIP W. HEDRICK*†,
STEVEN T. KALINOWSKI†§, DANIEL GARRIGAN† & WALTER M. BOYCE¶

†*Department of Biology, Arizona State University, Tempe, Arizona 85287 U.S.A. and* ¶*Department of Veterinary Pathology, Microbiology, and Immunology, University of California, Davis, California 95616 U.S.A.*

Bighorn sheep populations have greatly declined in numbers and distribution since European settlement, primarily because of high susceptibility to infectious diseases transmitted to them from domestic livestock. It has been suggested that low variation at major histocompatibility complex (MHC) genes, the most important genetic aspect of the vertebrate immune system, may result in high susceptibility to infectious disease. Therefore, we examined genetic polymorphism at a MHC gene (*Ovca-DRB*) in a large sample, both numerically and geographically, of bighorn sheep. Strikingly, there were 21 different alleles that showed extensive nucleotide and amino acid sequence divergence. In other words, low MHC variation does not appear to be the basis of the high disease susceptibility and decline in bighorn sheep. On the other hand, analysis of the pattern of the MHC polymorphism suggested that nonsynonymous substitutions predominated, especially at amino acids in the antigen-binding site. The average overall heterozygosity for the 16 amino acid positions that are part of the antigen binding site is 0.389 whereas that for the 67 amino acid positions not involved with antigen binding is 0.076. These findings imply that the diversity present in this gene is functionally significant and is, or has been, maintained by balancing selection. To examine the evolution of *DRB* alleles in related species, a phylogenetic analysis including other published ruminant (Bovidae and Cervidae) species, was carried out. An intermixture of sequences from bighorn sheep, domestic sheep, goats, cattle, bison, and musk ox was observed supporting trans-species polymorphism for these species. To reconcile the species and gene trees for the 104 sequences examined, 95 'deep coalescent' events were necessary, illustrating the importance of balancing selection maintaining variation over speciation events.

Keywords: adaptive variation, heterozygosity, phylogenetic tree, substitution rate, trans-species polymorphism.

Introduction

Bighorn sheep (*Ovis canadensis*) were abundant throughout much of western North American States before European settlement and Seton (1929), using distribution and density information, estimated that they numbered over a million animals. However, Buechner (1960), in reviewing the status of the species throughout its range, estimated only 25 000 animals

remained, a 40-fold reduction from the pre-settlement estimate. Disease transmission from livestock — domestic sheep (thought to be descended from the Old World sheep species, mouflon), goats, and cattle — was identified as the most important cause of this drastic reduction in numbers and distribution (Buechner, 1960). Many different types of diseases, including mite-induced scabies, bacterial pneumonia, and virus-caused blue tongue, have been implicated (Brown, 1993; Elliot *et al.*, 1994; Bunch *et al.*, 1999). In addition, other factors, such as overhunting, habitat loss and modification, competition for food and water from livestock, and predation, have also been suggested as significant. Bighorn sheep populations continued to decline until conservation and management policies focused on isolating bighorn sheep

*Correspondence. E-mail: philip.hedrick@asu.edu
‡Present address: Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica.
§Present address: Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Blvd. East, Seattle, Washington 98112, U.S.A.

439

from domestic livestock and began translocating animals back into historic ranges, which were now generally free of large numbers of livestock.

The vertebrate immune system has evolved into a complex physiological structure capable of recognizing foreign antigens (Janeway *et al.*, 1999). The most important genetic component of the mammalian immune system is a group of closely linked genes known as the major histocompatibility complex (MHC) which play a fundamental role in the immune response against pathogens (Hill, 1996; Hedrick & Kim, 2000). The MHC recognizes short peptides from pathogens and presents them to T cells to initiate the destruction of the pathogen. For example, in feral domestic sheep, MHC variants appear to play a major role in protection against strongyle nematode invasion, the most prevalent gastrointestinal parasite found (Paterson *et al.*, 1998) and in domestic sheep they are associated with the level of natural parasitic infection (Buitkamp *et al.*, 1996). In humans, the association of MHC polymorphism with resistance to infectious diseases, such as papilloma virus-induced cervical carcinoma, malaria, tuberculosis, hepatitis, and AIDS, has been shown (Hill *et al.*, 1991; Allen *et al.*, 1996; Hill, 1996; Thurz *et al.*, 1997; Carrington *et al.*, 1999). Because of this connection between MHC variability and the ability of the immune system to respond to a higher variety of pathogens, O'Brien & Evermann (1988) suggested that species or populations with low MHC diversity might be particularly vulnerable to infectious disease.

Boyce *et al.* (1997), using RFLP analyses, examined MHC variation in bighorn sheep from some of the same populations in this study. However, because of the nature of the molecular techniques they used, only a gross-level examination was provided for the variation present at the MHC loci, particularly because the positions of the recognition sites for the RFLPs were not known. Therefore, sequencing of a translated segment of a MHC gene, as we have done here, provides a much more direct approach to studying functional polymorphism within the MHC region.

Specifically, here we examine the amount of variation for a class II (genes involved in recognizing extracellular antigens, including those from bacteria) MHC gene, *DRB*. This gene was chosen because it is the most polymorphic class II gene in other species (e.g. humans and cattle) and much of this variation is in the functionally important antigen binding site (ABS). If the high susceptibility of bighorn sheep to infectious disease were the result of low MHC variation, then only a few, very similar alleles and low heterozygosities would be predicted. However, as we show below, there is extensive MHC variation for the *DRB* gene, suggesting that low MHC variation does not appear to be

the cause of decline of bighorn sheep from infectious disease. In fact, MHC variation over populations shows a similar pattern to that observed in a study of microsatellite loci in the same individuals. On the other hand, we show that the pattern of MHC sequence variation in bighorn sheep is consistent with balancing selection.

Materials and methods

Samples

We examined samples from 206 bighorn sheep from 11 desert bighorn populations and seven individuals from a population of Rocky Mountain bighorn sheep, *O. c. canadensis* (Fig. 1). In the desert bighorn samples, we have samples from three of the putative desert bighorn sheep subspecies, *O. c. cremnobates*, *O. c. mexicana*, and *O. c. nelsoni*. Arizona Game & Fish Department personnel and hunters collected 95 blood samples from Arizona bighorn sheep and W. Boyce provided DNA from 118 animals from California and New Mexico. These study sites are composed of native bighorn sheep populations except for the Stewart Mountain, Red Rock Refuge and Wheeler Peak study sites. The Stewart Mountain individuals descended from sheep transplanted from the Kofa Mountains of Arizona; the ancestors of the Red Rock Refuge bighorn sheep were captured in the San Andres Mountains of New Mexico and Loma Prieta of Sonora, Mexico; and Wheeler Peak Rocky Mountain animals descended from sheep transplanted from Banff, Alberta, Canada.

Molecular techniques

Details of the molecular techniques are described in Gutierrez-Espeleta (1999). We examined variation for a highly polymorphic 249-nucleotide sequence of exon 2 of the class II *DRB* gene. PCR amplification of this region was carried out using primers LA31 and LA32 of Sigurdardottir *et al.* (1991). To identify allelic diversity, all individuals were subjected to single strand conformational polymorphism analysis (SSCP) (Orita *et al.*, 1989). All unique conformations were cloned and subclones were screened by PCR to ensure that the products contained the correct length insert. Approximately 20 subclones were screened by SSCP to ensure that only the subclones with SSCP banding patterns corresponding to the original genomic sample were sequenced. At least two subclones of each individual allele per animal (two for homozygotes and four for heterozygotes) were sequenced. Sequences were confirmed by comparing the two sequences from the

Fig. 1 Location of natural desert bighorn sheep populations in Arizona and California. Not shown are the locations of Stewart Mountain, AZ (a transplant population from Kofa Mountains, AZ), Red Rock, NM (a captive population primarily from San Andres Mountains, NM), and Wheeler Peak, NM (a transplant population from Banff National Park, Alberta, Canada).
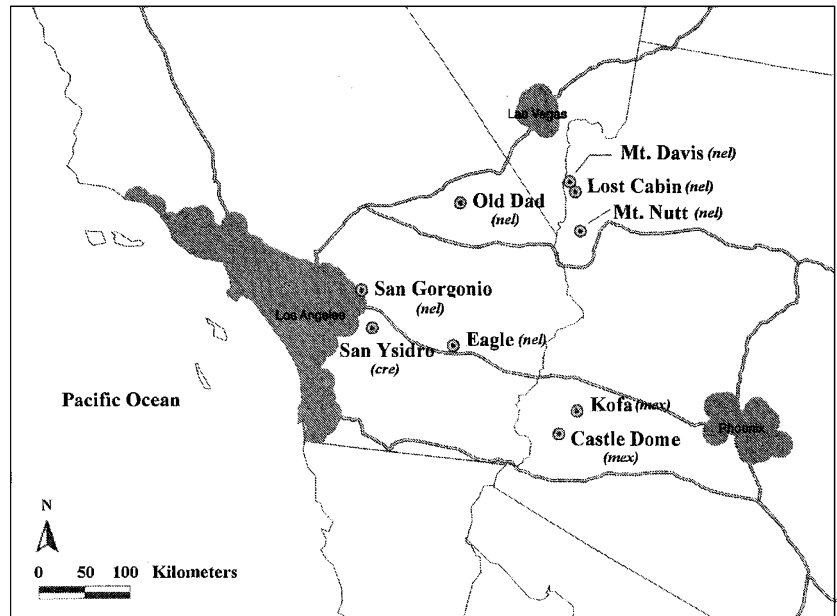
same allele and by sequencing the same allele from more than one individual. Some of the early sequence reactions were performed manually but most used an Applied Biosystems 373 A automated sequencer. The 21 *Ovca-DRB* sequences identified in this study have been deposited in GenBank under accession numbers AF324840–AF324861.

Data analysis

MHC sequence alignments were performed using the EYEBALL sequence editor v. 2.00. A neighbour-joining tree of all nucleotide substitutions using the correction of Jukes & Cantor (1969) for multiple hits was constructed using the MEGA v. 1.01 program (Kumar *et al.*, 1993), for the 21 *Ovca-DRB* alleles and 122 *DRB* alleles from ruminant species found on the GenBank database (Bovidae: 22 cattle, 11 bison, 1 musk ox, 16 goat, 33 domestic sheep; Cervidae: 10 moose, 3 roe deer, 9 reindeer, 2 fallow deer, and 15 white-tailed deer), and bootstrapped (1000 replicates) to give a consensus tree with confidence limits on branch points (data not shown). Because all *Ovca-DRB* alleles were contained within the Bovidae lineage, a similar neighbour-joining tree was constructed for the 104 sequences in the Bovidae lineage. We then used the method of reconciled phylogenetic trees (Page & Charleston, 1997) that finds the number of 'deep coalescent' events necessary to make the gene tree concordant with species tree.

MEGA was used to estimate the rate of nonsynonymous (d_N) and synonymous (d_S) substitutions according to Nei & Gojobori (1986), applying the correction of Jukes & Cantor (1969) for multiple hits. The difference between these rates was evaluated with a t-test with infinite degrees of freedom according to the test statistic $t = d/s(d)$, where $s(d)$ is the standard error of d and is given by $s(d) = [\mathrm{Var}(d_S) + \mathrm{Var}(d_N)]^{1/2}$ (Kumar *et al.*, 1993). We also calculated the heterozygosity for each polymorphic amino acid position, either using the average weighted allelic frequency over all samples to get H_T or calculating the heterozygosity within each sample and averaging this over all samples, H_S (e.g. Nei, 1987).

In addition to these sequence-based measures of diversity and differentiation we analysed the frequencies of MHC alleles with standard population genetic measures. Expected heterozygosity (gene diversity) (e.g. Nei, 1987) was used as a measure of genetic variation within each population. Deviations from Hardy–Weinberg proportions were tested by using GENEPOP 3.1d (Raymond & Rousset, 1997). This same program was used to calculate F_{ST} values to measure the extent of divergence over populations. The genetic distance of Nei (D_A) (Nei, 1987) was used to determine differentiation between each pair of study sites and regional groupings of populations. D_A was deemed an appropriate measure because a number of pairs of populations did not share any alleles so that the standard genetic distance of Nei (1987) becomes infinity. D_A was used to obtain a neighbour-joining dendrogram for the 12 bighorn sheep populations using the MEGA v. 1.01 program (Kumar *et al.*, 1993) and to compare genetic distance with geographical distance between the samples.

Results

Genetic variation within population

SSCP and nucleotide sequence analyses revealed 18 novel *Ovca-DRB* exon 2 allele sequences among the 206 desert bighorn sheep and three unique *Ovca-DRB* alleles in seven Rocky Mountain bighorn sheep. All populations had high amounts of genetic variation as shown by the number of alleles, ranging from 3 to 8 per population (mean of 5.5) (Table 1). Also the heterozygosity was quite high with the Kofa Mountains population having the highest expected heterozygosity ($H_E = 0.964$), the San Ysidro population the lowest one ($H_E = 0.489$), and the average heterozygosity was 0.742 (Table 1). Four of the populations (Mt. Nutt, Castle Dome Mountains, Stewart Mountains, Red Rock Refuge) had a significant ($P < 0.05$) deficiency of heterozygotes compared to Hardy–Weinberg proportions. The same individuals did not show a deficiency of heterozygotes for 10 microsatellite loci (Gutierrez-Espeleta *et al.*, 2000), suggesting that a Wahlund effect (Hedrick, 2000) is unlikely. Another possible explanation for this deficiency is the presence of a null allele. Therefore, we calculated the frequency of putative null allele for all populations with a deficiency of heterozygotes using the approach of Brookfield (1996) (Table 1). The estimated frequency of the null allele in these four populations ranged from 0.066 to 0.120. Because we did not actually observe any null homozygotes and these estimates are not high for recessive null alleles, in our analyses below we did not include the putative null allele.

Genetic variation between populations

We estimated F_{ST} values for different regional groupings of populations. Genetic differentiation among the three northern Arizona populations and among the three southern Arizona populations was low ($F_{ST} = 0.041$ and 0.000, respectively), whereas it was higher among the six Arizona populations ($F_{ST} = 0.084$). The estimate of F_{ST} for all desert populations was 0.213, and when all 12 populations (desert and Rocky Mountain bighorn sheep) were combined, the F_{ST} was 0.221.

We also made comparisons among regional grouping of populations using D_A values. When we compared distances between northern Arizona samples, D_A averaged 0.155, whereas between southern Arizona samples, D_A averaged 0.271. When northern and southern Arizona samples were compared, the average pairwise D_A value was 0.402. We use pairwise D_A values to build the dendrogram depicted in Fig. 2, which provides one method of summarizing the genetic relationships between study sites. In general, the northern Arizona *O. c. nelsoni*

Table 1 Location, putative subspecies of bighorn sheep, number of individuals sampled (N), number of alleles (n), expected (H_E) and observed (H_O) heterozygosity, and estimated null allele frequency (p_n) for the four locations with a significant heterozygous deficiency

Location	Putative subspecies	N	n	H_E	H_O	p_n
Northern Arizona						
Mt. Davis	*O. c. nelsoni*	15	6	0.736	0.600	—
Lost Cabin	*O. c. nelsoni*	15	4	0.752	0.667	—
Mt. Nutt	*O. c. nelsoni*	27	5	0.766	0.556*	0.119
Southern Arizona						
Kofa Mountains	*O. c. mexicana*	4	7	0.964	0.750	—
Stewart Mountain	*O. c. mexicana*	14	8	0.868	0.643*	0.120
Castle Dome Mountains	*O. c. mexicana*	20	7	0.850	0.700*	0.081
Southern California						
Old Dad Mountains	*O. c. nelsoni*	23	5	0.745	0.652	—
Eagle Mountains	*O. c. nelsoni*	21	9	0.867	0.810	—
San Gorgonio	*O. c. nelsoni*	20	3	0.595	0.600	—
San Ysidro	*O. c. cremnobates*	22	6	0.489	0.591	—
New Mexico						
Red Rock Refuge	*O. c. mexicana*	25	3	0.659	0.560*	0.066
Rocky Mountains						
Wheeler Peak, NM	*O. c. canadensis*	7	3	0.616	0.429	—
Total or mean		213	5.5	0.742	0.630	—

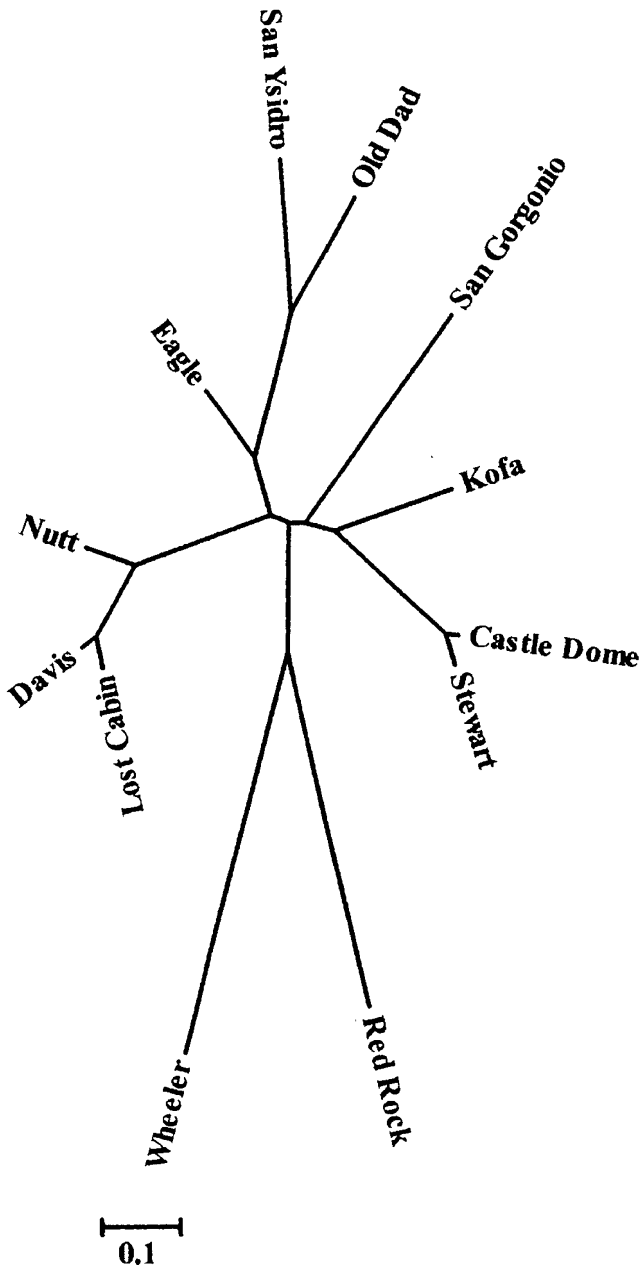*H_O is significantly lower than H_E ($P < 0.05$).

Fig. 2 Neighbour-joining tree based on D_A genetic distance for the 12 bighorn sheep populations included in this study.

populations, Mt. Davis, Lost Cabin, and Mt. Nutt, cluster together, and the southern Arizona *O. c. mexicana* populations, Kofa, Castle Dome, and Stewart Mountains, cluster together consistent. Wheeler (Rocky Mountain bighorn) and Red Rock (desert bighorn) populations appeared to be clear outgroups relative to the other desert populations. Besides these observations, there is little significant clustering of other taxa.

Genetic distance generally increases over time and over geographical distance for neutral microsatellite markers (Gutierrez-Espeleta *et al.*, 2000). To determine
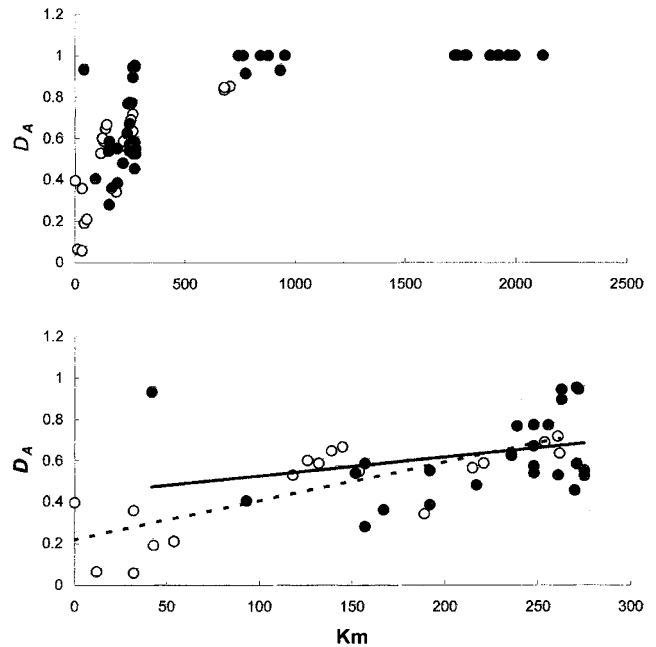

Fig. 3 Pairwise genetic distances (D_A) plotted against geographical distance (km) for all comparisons (upper graph) and for comparisons < 300 km apart (lower graph), in bighorn sheep populations. Comparisons within and between populations from putative subspecies are indicated by open and closed circles, respectively. The linear regression slopes are given for the populations < 300 km apart for the within (broken line) and between (solid line) putative subspecies in the lower part of the figure.

if this relationship is also true for this MHC gene, we plotted genetic and geographical distance for all pairs of populations (Fig. 3). For nearly all the pairs of populations that are separated by more than 300 km, D_A is near its maximum of 1. However, for the pairs of populations less than 300 km apart, genetic distance increases significantly with geographical distance. In the lower part of Fig. 3, this linear regression is given separately for within putative subspecies comparisons (open circles) and between putative subspecies comparisons (closed circles). These are both significantly different from zero but are not significantly different from each other. This pattern of genetic distance over space is similar to that observed for 10 microsatellite loci in the same populations (Gutierrez-Espeleta *et al.*, 2000) and consistent with the hypothesis that non-selective factors are of primary importance in determining spatial differentiation at this locus.

Sequence analysis

Analysis of the 21 alleles showed that 22% (56 out of 249) of the nucleotides and 40% (33 out of 83) of the

Table 2 Amino acid sequence at the variable codon positions for the 21 *Ovca-DRB* alleles in bighorn sheep, omitting the duplication in *Ovca-DRB**13. Identity to the consensus sequence is indicated by – and ABS positions by *

Position	11*	13*	16	18	26	28*	30*	31	32	37*	38*	42	47	50	51	52	53	55	56	57	59	60	61*	66	67*	70*	71*	74*	77	78*	81	86*	90
Consensus	H	S	R	F	F	D	Y	F	Y	N	V	S	Y	V	A	E	L	R	P	D	K	Y	W	E	L	R	R	E	T	Y	H	F	T
Ovca-DRB*1	Y	G	H	–	L	H	F	Y	T	T	–	–	F	–	T	Q	Q	Q	E	–	E	H	–	–	I	Q	K	–	–	V	–	–	–
Ovca-DRB*2	Y	G	H	–	L	H	F	Y	T	T	–	–	F	–	T	Q	Q	Q	E	–	E	H	–	–	I	Q	K	–	–	V	–	–	–
Ovca-DRB*3	Y	G	H	–	L	H	F	Y	T	T	–	–	F	–	T	Q	Q	Q	E	–	E	H	–	–	I	Q	K	–	–	V	–	V	–
Ovca-DRB*4	Y	G	H	–	L	H	F	Y		T	–	–		–	T	–	–	–	E	–	E	H	–	–	I	Q	K	–	–	V	–	–	I
Ovca-DRB*5	T	K	–	S	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	G	–
Ovca-DRB*6	T	K	–	S	–	–	–	–	–	Y	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	G	–
Ovca-DRB*7	T	K	–	S	–	–	–	–	–	Y	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	I	–
Ovca-DRB*8	T	K	–	S	–	–	–	–	–	–	–	–	–	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	G	–
Ovca-DRB*9	–	–	–	–	–	–	–	–	–	Y	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	N	–	–	–	G	–
Ovca-DRB*10	–	–	H	S	L	–	–	–	H	F	–	–	F	–	–	–	–	–	–	A	E	–	Y	D	F	S	–	A	–	–	R	I	S
Ovca-DRB*11	–	–	–	S	–	–	–	–	–	Y	A	–	–	–	–	–	–	–	R	S	E	–	–	–	–	–	–	–	–	–	–	I	–
Ovca-DRB*12	–	–	–	S	Y	–	–	–	H	F	–	–	–	–	–	–	–	–	–	–	–	–	–	–	I	–	K	N	–	–	–	–	S
Ovca-DRB*13	–	–	–	S	Y	–	–	–	H	F	–	–	–	–	–	–	–	–	–	–	–	–	–	–	I	–	K	N	–	–	–	–	S
Ovca-DRB*14	–	–	H	S	L	–	–	–	H	F	–	–	F	–	–	–	–	–	–	A	E	–	Y	D	F	S	T	A	–	–	–	I	S
Ovca-DRB*15	–	–	–	–	Y	–	–	–	–	–	–	N	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	S
Ovca-DRB*16	R	–	–	–	–	–	–	–	–	Y	A	–	–	–	–	–	–	–	R	S	E	–	–	–	–	–	–	–	–	–	–	G	–
Ovca-DRB*17	A	–	H	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Q	T	–	–	–	–	–	S
Ovca-DRB*18	A	–	–	–	E	–	–	–	–	Y	–	N	–	–	–	–	–	–	–	–	–	–	–	–	I	–	K	N	–	–	–	G	–
Ovca-DRB*19	A	–	H	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	D	F	–	K	A	–	–	–	–	S
Ovca-DRB*20	A	–	H	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	K	N	–	–	–	–	S
Ovca-DRB*21	A	–	H	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Q	T	–	R	–	–	–	S

amino acid positions (Table 2) were variable. Nucleotide sequence variation among all pairwise comparisons of *Ovca-DRB* alleles ranged from 1 to 36 nucleotides, whereas amino acid replacements among all pairwise comparisons ranged from 0 (only synonymous substitutions, i.e. alleles *Ovca-DRB*1 and *Ovca-DRB*2) to 25. The most common allele *Ovca-DRB*7 was present in nine of the 11 desert bighorn populations and had an average weighted allelic frequency of 0.141 (Appendix). The rarest alleles in desert bighorn were *Ovca-DRB*8 and *Ovca-DRB*15 which were found as heterozygotes in one animal each. Five of the alleles were unique to a single desert bighorn population and all three alleles found in the Rocky Mountain sample were not found in the desert bighorn samples. Alleles *Ovca-DRB*12 and *Ovca-DRB*13 were identical, except that the latter contained a tandem duplication of four nucleotides (CGGG) starting at amino acid position 85, which produced a stop codon at position 88. The *Ovca-DRB*13 allele was only identified as a heterozygote in two individuals in the Stewart Mountains population.

Fourteen of the 16 (88%) antigen binding site (ABS) positions had nonsynonymous substitutions while 19 of the 67 (28%) of the non-ABS positions had nonsynonymous substitutions (Table 2). One ABS position (11) had five different amino acids over the 21 alleles and two others (37 and 86) had four different amino acids. The rate of nonsynonymous substitutions (d_N) occurred at a significantly higher frequency than that of synonymous substitutions (d_S) (Table 3). For the ABS positions, the ratio of the estimated rates of nonsynonymous to synonymous substitutions was 7.4. Five positions with only synonymous substitutions were found scattered throughout the sequence and none of them were in the ABS.

The frequencies of the different sequences can be combined in a graph of the heterozygosity, where H_S and H_T are the average heterozygosity within populations and over the total sample, respectively, for the 33 polymorphic amino acid positions (Fig. 4). The highest values are for position 11 where H_S and H_T are 0.60 and 0.77, respectively. The average value of H_T for the ABS sites (indicated by *) is 0.389 and for the non-ABS sites it is 0.076, a fivefold higher value for the ABS sites. Most of the variation is within population, with H_S/H_T averaging about 0.8 over the individual amino acids.
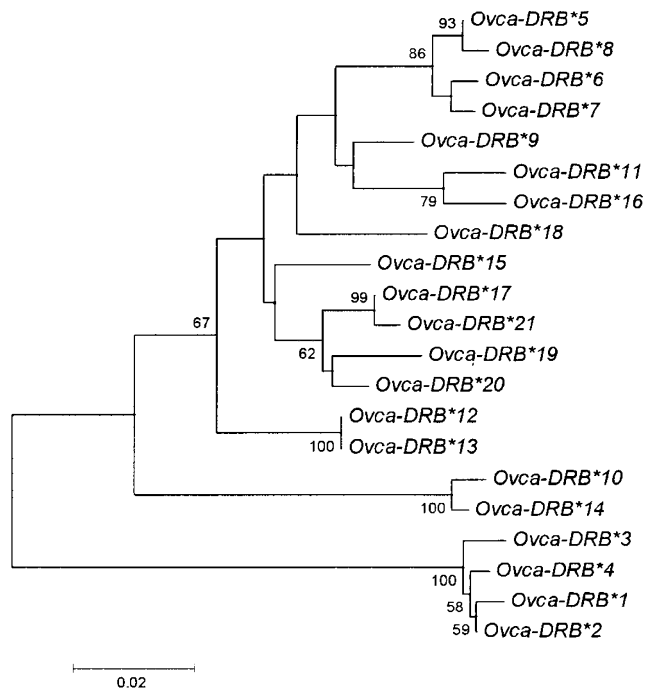
First, Fig. 5 gives a neighbour-joining tree of the 21 bighorn sheep *DRB* alleles. A number of allele lineages were very strongly supported, for example, the two lineages of alleles *Ovca-DRB*1–4 and *Ovca-DRB*10 and 14, both of which had a number of differences from the other sequences, clustered with bootstrap values

Table 3 Rates of nonsynonymous (d_N) and synonymous (d_S) substitutions and their ratio N is the number of codons and P indicates probability of significance between d_N and d_S

| Positions	N	d_N	d_S	d_N/d_S	P
Antigen binding	16	0.350 ± 0.059	0.047 ± 0.044	7.45	< 0.0005
Non-antigen binding	67	0.052 ± 0.011	0.036 ± 0.017	1.45	> 0.2
All	83	0.100 ± 0.013	0.038 ± 0.016	2.63	< 0.005
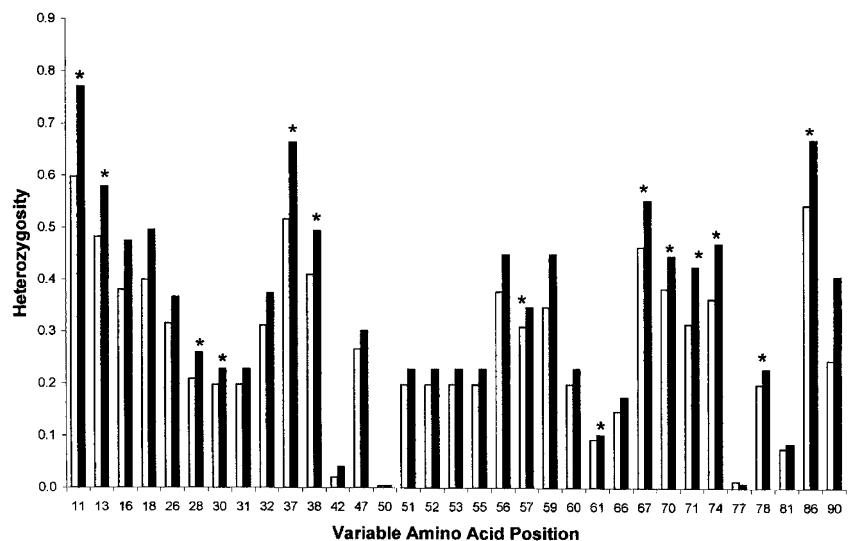
Fig. 4 The average heterozygosity of the polymorphic amino acid positions within populations of bighorn sheep, H_S (open bar), and over all the populations, H_T (closed bar). *Indicates antigen-binding positions.

Fig. 5 Neighbour-joining tree of the sequences of the *DRB* alleles found in bighorn sheep. The numbers indicate bootstrap significance values (1000 replicates) and the scale indicates Jukes and Cantor corrected genetic distance.

of 100%. Another lineage of four sequences, *Ovca-DRB**5–8, clustered together with a bootstrap level of 86%. The three Rocky Mountain alleles, *Ovca-DRB**4, 18, and 21, were spread around the tree and were 1, 7, and 1 amino acids (1, 8, and 1 nucleotides) different from the closest desert bighorn sheep allele. Overall, most of the internal branches did not have strong bootstrap support whereas a number of terminal groups showed significant bootstrap clustering.

Second, we examined the *DRB* sequences from bighorn sheep and 10 other ruminant species. These sequences resulted in two clear clusters; one corresponded to the Bovidae lineage (domestic and bighorn sheep, musk ox, goat, cattle and bison) and the other one to the Cervidae lineage (moose, roe deer, reindeer, fallow deer and white-tailed deer) (see also Mikko *et al.*, 1999). Therefore, another neighbour-joining tree for the *DRB* sequences was constructed for just the Bovidae lineage alleles (right side of Fig. 6) and to show how these correspond to the species tree (left side of Fig. 6). For neutral loci, species and gene trees are generally concordant except for genes that are polymorphic at the time of speciation (Nei & Kumar, 2000). Unlike the expected pattern for neutral loci, here the bighorn sheep alleles (and sheep, goat, bison, and cow alleles) are intermixed with sequences from different species. In addition, within species there are clusters of closely
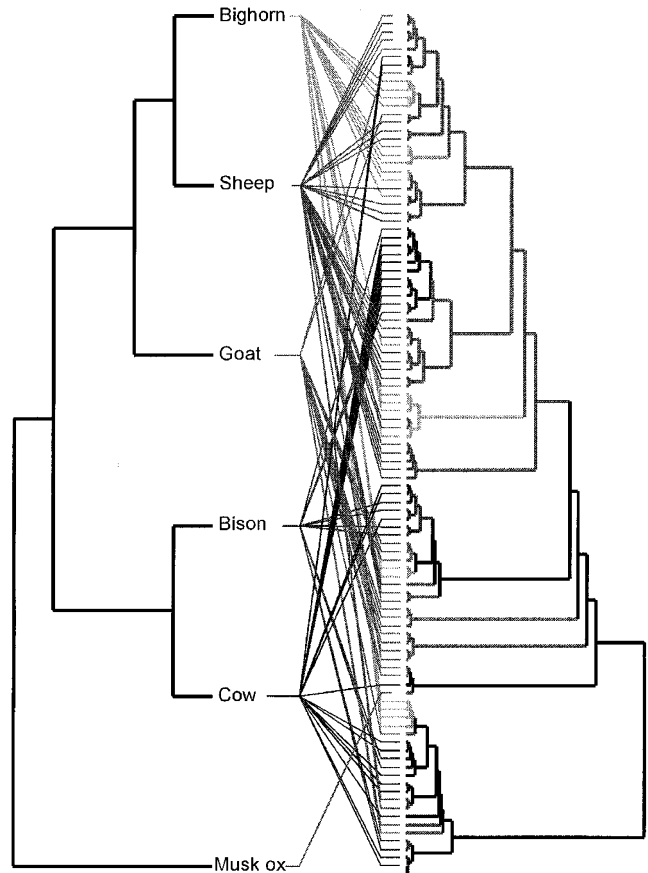
Fig. 6 A neighbour-joining gene tree for 104 *DRB* sequences from six Bovidae species (right) reconciled with true species tree (left) by the method of Page & Charleston (1997). Separation of lines radiating from the tips of the species tree into different branches of the gene tree indicates extensive 'trans-species' evolution of *DRB* alleles.

related sequences, or lineages. Also note that the bighorn sheep sequences are intermingled with the domestic sheep alleles, although none of the bighorn and domestic sheep alleles are identical.

To quantify the number of differences between the species tree and the *DRB* gene tree, we used the method of reconciled phylogenetic trees (Page & Charleston, 1997) that finds the number of 'deep coalescent' events necessary to make the gene and species trees concordant for the 104 *DRB* sequences in Fig. 6. A 'deep coalescent' event, in this case, can be defined as polymorphism in the ancestral species being retained in each of the descendant species, a phenomenon also known as trans-species evolution (Klein *et al.*, 1998). If all *DRB* sequences diversified after the origin of each species, the lines radiating from the tips of the species tree would cluster in a single clade on the gene tree. To account for the large discrepancies between the species and gene trees here, we need to invoke 95 'deep coalescent' events

to reconcile the two trees. Of particular interest in Fig. 6 is the observation that the most basal branch of the gene tree is comprised of sequences from each species, save the one musk ox sequence. This indicates that this particular ancient lineage of alleles has been preserved in each descendant species by balancing selection. Although we have not estimated the age of these *DRB* allelic lineages, it is clear that most major lineages in the gene tree diverged well before the family level diversification of the Artiodactyla.

Discussion

Even though bighorn sheep populations have declined 40-fold between the time of European settlement and today and many local populations have gone extinct, primarily because of mortality due to infectious disease, we found extensive genetic variation at an important MHC gene. As a result, it does not appear that low MHC variation was the source of the high susceptibility to infectious disease observed as generally suggested by O'Brien & Evermann (1988). In fact, one might conclude from our study that high genetic variation at an MHC gene is no safeguard against the detrimental effects of infectious disease. In addition, Mikko *et al.* (1999) have found that beaver, moose, and roe deer, all of which have low MHC variation, showed dramatic population expansions in Sweden in the last century. Therefore, a strong association of the extent of MHC variation and disease resistance does not appear universal and other, probably nongenetic, factors appear to be of overriding significance in determining dramatic changes in population numbers.

In the bighorn sheep examined we identified a total of 21 *Ovca-DRB* alleles, a number greater than the number of *DRB* alleles detected in some other ruminant species populations i.e. musk ox, moose, roe deer, reindeer (Mikko & Andersson, 1995; Mikko, 1997); bison (Mikko *et al.*, 1997); white-tailed deer (Van Den Bussche *et al.*, 1999) and similar to that at the most polymorphic MHC class II locus in cattle (Davies *et al.*, 1997), red deer (Swarbrick *et al.*, 1995), goat and domestic sheep (Schwaiger *et al.*, 1994). Our study revealed high levels of MHC-*DRB* genetic variation in bighorn sheep as shown both by the number of alleles and heterozygosity. Nucleotide and amino acid sequence variation at positions associated with the antigen binding site were particularly high. For example, at one ABS amino acid position, we found five different amino acids over the 21 alleles and an overall heterozygosity of 0.77. These high amounts of variation, consistent with our previous microsatellite data (Gutierrez-Espeleta *et al.*, 2000), indicate that there is a large reservoir of nuclear DNA variation in desert

bighorn sheep populations across the south-western United States.

However, we have only studied one MHC gene, out of many genes involved in the immune response in bighorn sheep, and that other MHC genes could potentially show other patterns within and between populations. As context, the MHC in humans has 128 genes predicted to be expressed and about 40% (around 50 genes) are thought to be involved with the immune response (Beck & Trowsdale, 2000). However, the *DRB* gene that we have investigated is one of only a few class II genes expressed in the Bovidae and is homologous to the class II gene fundamental to transplant success and disease resistance in humans. Further, the class II region genes are closely linked in humans (Marsh *et al.*, 2000), and all other mammals examined, and variants at these genes are generally in strong linkage disequilibrium. In other words, the patterns that we have observed for *DRB* should be a good indicator of the genetic variation in other class II genes and even some other less closely linked genes in the MHC.

Although we cannot implicate low MHC variation in bighorn sheep in their numerical decline, there are three lines of evidence that indicate a role for natural selection in the persistence and diversity of MHC alleles in our data. First, the observed excess of nonsynonymous over synonymous substitutions among the 21 sequences, especially at antigen-binding sites, indicates that nonsynonymous sites evolve faster than synonymous sites and imply the action of balancing selection (or positive Darwinian selection) to favour new variants and increase MHC diversity (Hughes & Nei, 1988, 1989). This pattern has since been found in a number of other species, but particularly noteworthy here is a similar result found in feral domestic sheep (Paterson, 1998). Second, when the frequencies of the 21 alleles are taken into account, the heterozygosity at the antigen-binding sites is 0.389, compared to 0.076 for the non-antigen-binding sites. This fivefold higher heterozygosity at functionally important positions compared to those with no known ABS function, many of which are at adjacent amino acid positions, strongly implies that the extent of variation is associated with function for this MHC molecule (see also Hedrick *et al.*, 1991).

Third, when MHC sequences from different taxa are compared, it is frequently observed that many allelic lineages appear older than the species from which they were detected (Figueroa *et al.*, 1988; Klein *et al.*, 1993). This phenomenon, known as trans-species polymorphism, implies maintenance of MHC diversity by balancing selection over long periods of evolutionary time. Some bighorn allele lineages were more similar to goat or domestic sheep alleles than to other bighorn allele

lineages, supporting the trans-species persistence of *DRB* allelic lineages. *Ovca-DRB* allele lineages were distributed throughout the tree in the Bovidae lineage on the basis of sequence similarity and all species (bighorn sheep, domestic sheep, cattle, bison, goat and musk ox) alleles were intermingled, that is, there was no clustering of all alleles within a species of these MHC sequences. The finding that a very large number of deep coalescent events are necessary to reconcile the gene tree with the species tree suggests that balancing selection has played an overriding role in the evolution of *DRB* in these species and their ancestors.

Comparison of MHC and microsatellite variation

Examining and comparing genetic variation for MHC and microsatellite loci in the same individuals provides an opportunity to evaluate the neutral and selective forces influencing genetic variation, although this needs to be done very cautiously. Here we can compare the data from our previous study in the same populations, and many of the same individuals, of 10 highly variable microsatellite loci (Gutierrez-Espeleta *et al.*, 2000) with the MHC variation documented in this study. First, we found that the amount of expected heterozygosity in these populations was always greater or equal for the MHC locus than for the average of the 10 microsatellite loci (the difference ranged from 0.00 to 0.36). However, we used different genetic distance measures (D_S for the microsatellite loci and D_A for the MHC locus) and we used an unbiased measure of D_S while there is no comparable, unbiased measure for D_A. In addition, the microsatellite loci may more closely fit the stepwise-mutation model whereas the MHC locus may more closely fit the infinite-allele model. Further, the mutation rate for microsatellite loci may be higher than that for the MHC locus although a high gene conversion rate has been found for a human class II MHC gene (for a review, see Hedrick & Kim, 2000). These differences suggest that any comparison of levels of variation between the two types of loci should only be very general.

If balancing selection in the form of heterozygote advantage acted in the present generation, this effect appears to be small enough to not increase the number of heterozygotes over Hardy–Weinberg expectations, as has been observed in some studies (Black & Salzano, 1983; Markow *et al.*, 1993). However, the possible presence of null alleles and the small sample sizes in some of the populations might make it difficult to detect the contemporary impact of selection on genotypic proportions.

Second, to determine if there was an association in the genetic variation for microsatellite and MHC loci over these populations, we calculated the correlation for expected heterozygosity and observed number of alleles. There was a strong positive association, 0.56 ($P < 0.10$) for heterozygosity and 0.67 ($P < 0.05$) for the number of alleles. Because molecular analysis of MHC variation indicates selective maintenance, as discussed above, this correlation with neutral variation appears paradoxical. However, a possible explanation is that the average selection coefficient per generation against MHC homozygotes is only around 0.02 (Satta *et al.*, 1994). This may be large enough to result in the molecular selective patterns but too small to overcome the impact of non-selective factors influencing differences in the amount of genetic variation observed in the populations examined. A similar strong association of MHC and microsatellite variation has been documented for Arizona populations of the endangered Gila topminnow, a species recognized to undergo large variations in population size (Hedrick *et al.*, 2001). In addition, over different species the level of MHC variation appears correlated with genomic diversity in general (Edwards & Potts, 1996; Mikko *et al.*, 1999). These associations would obviously make it difficult to attribute differential mortality over populations or species to MHC or other loci because the pattern of variation for different loci often strongly covary.

Third, variation over populations at both MHC and microsatellite loci should be influenced by non-selective forces, but if selective factors (i.e. pathogens) acting on MHC have been of significantly greater magnitude than the non-selective factors, then the spatial patterns of variation for MHC and microsatellite loci may differ. Here mean F_{ST} values were very similar for both markers for comparisons within and across regions, but in all cases, MHC F_{ST} values were slightly lower (0.03–0.08). This difference may be related to the amount of genetic variation at the respective loci because the maximum value F_{ST} can have is dependent on the amount of variation in the population, i.e. markers with lower heterozygosity would be expected to have lower F_{ST} values (Hedrick, 1999). For the populations examined for both markers, the average within-population heterozygosity for the MHC locus was 0.742 and that for microsatellite loci was 0.516, a difference that could produce F_{ST} values consistent with those observed.

Acknowledgements

References

ALLEN, M., KALANTARI, M., YLITARO, N., PETTERSON, B. ET AL. 1996. HLA DQ-DR haplotype and susceptibility to cervical carcinoma: indications of increased risk for development of cervical carcinoma in individuals infected with HPV 18. *Tissue Antigens*, **48**, 32–37.

BECK, S. AND TROWSDALE, J. 2000. The human Major Histocompatibility Complex: lessons from the DNA sequence. *Ann. Rev. Genomics Hum. Genet.*, **1**, 117–137.

BLACK, F. L. AND SALZANO, F. M. 1983. Evidence for heterosis in the HLA system. *Am. J. Hum. Genet.*, **33**, 894–899.

BOYCE, W. M., HEDRICK, P. W., MUGGLI-COCKETT, N. E., KALINOWSKI, S. ET AL. 1997. Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics*, **145**, 421–433.

BROOKFIELD, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol. Ecol.*, **5**, 453–455.

BROWN, D. 1993. Early history. In: Lee, R. (ed.) *The Desert Bighorn Sheep in Arizona*, pp. 1–11. Arizona Game and Fish Department, Phoenix.

BUECHNER, H. K. 1960. The bighorn sheep of the United States: its past, present and future. *Wildlife Monogr.*, **4**, 1–174.

BUITKAMP, J., FILMETHER, P., STEAR, M. J. AND EPPLEN, J. T. 1996. Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *Parasitol. Res.*, **82**, 693–696.

BUNCH, T. D., BOYCE, W. M., HILBER, C. P., LANCE, W. R. ET AL. 1999. Diseases of North American Wild Sheep. In: Valdez, R. and Krausman, P. R. (eds) *Mountain Sheep of North America*, pp. 209–237. University of Arizona Press, Tucson, AZ.

CARRINGTON, M., NELSON, G. W., MARTIN, M. P., KISSNER, T. ET AL. 1999. HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, **283**, 1748–1752.

DAVIES, C. J., ANDERSSON, L., ELLIS, S., HENSEN, E. J. ET AL. 1997. Nomenclature for Factors of the BoLA system. Report of the ISAG BoLA Nomenclature Committee.

EDWARDS, S. V. AND POTTS, W. K. 1996. Polymorphism of genes in the major histocompatibility complex (MHC): Implications for conservation genetics of vertebrates. In: Smith, T. B. and Wayne, R. K. (eds) *Molecular Genetic Approaches in Conservation*, pp. 214–237. Oxford University Press, New York.

ELLIOT, L. F., BOYCE, W. M., CLARK, R. K. AND JESSUP, D. A. 1994. Geographic analysis of pathogen exposure in bighorn sheep (*Ovis canadensis*). *J. Wildlife Dis.*, **30**, 315–318.

FIGUEROA, F., GUNTHER, E. AND KLEIN, J. 1988. MHC polymorphism pre-dating speciation. *Nature*, **335**, 265–267.

GUTIERREZ-ESPELETA, G. A. 1999. *Neutral and Adaptive Genetic Variation in Desert Bighorn Sheep* (Ovis canadensis): implications for Conservation. PhD Thesis, Arizona State University, Tempe.

GUTIERREZ-ESPELETA, G. A., KALINOWSKI, S. T., BOYCE, W. M. AND HEDRICK, P. W., 2000. Genetic variation and population structure in desert bighorn sheep: implications for conservation. *Cons. Genet.*, **1**, 3–15.

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

HEDRICK, P. W. 2000. *Genetics of Populations*, 2nd edn. Jones and Bartlett, Boston.

HEDRICK, P. W. AND KIM, T. 2000. Genetics of complex polymorphisms: parasites and maintenance of MHC variation. In: Singh, R. S. and Krimbas, C. B. (eds) *Evolutionary Genetics from Molecules to Morphology*, pp. 204–234. Cambridge University Press, New York.

HEDRICK, P. W., WHITTAM, T. S. AND PARHAM, P. 1991. Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 5897–5901.

HEDRICK, P. W., PARKER, K. M. AND LEE, R. N. 2001. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Mol. Ecol.*, in press.

HILL, A. V. S. 1996. Genetics of infectious disease resistance. *Curr. Opin. Genet. Develop.*, **6**, 348–353.

HILL, A. V. S., ALLSOPP, C. E. M., KWIATKOWSKI, D., ANSTEY, N. M. ET AL. 1991. Common West African HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.

HUGHES, A. L. AND NEI, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature*, **335**, 167–170.

HUGHES, A. L. AND NEI, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 958–962.

JANEWAY, C. A., TRAVER, P., WALPORT, M. AND CAPRA, J. D. 1999. *Immunobiology*, Garland Publishers, London.

JUKES, T. H. AND CANTOR, C. R. 1969. Evolution of protein molecules. In: Munroe, H. N. (ed.) *Mammalian Protein Metabolism*, vol. 3. Academic Press, New York.

KLEIN, J., SATTA, Y., TAKAHATA, N. AND O'HUIGIN, C. 1993. Trans-specific MHC polymorphism and the origin of species in primates. *J. Med. Primatol.*, **22**, 57–64.

KLEIN, J., SATO, A., NAGL, S. AND O'HUIGIN, C. 1998. Molecular trans-species polymorphism. *Ann. Rev. Ecol. Syst.*, **29**, 1–21.

KUMAR, S., TAMURA, K. AND NEI, M. 1993. MEGA: *Molecular evolutionary genetics analysis*, version 1.01. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA.

MARKOW, R., HEDRICK, P. W., ZUERLEIN, K., DANILOVS, J. ET AL. 1993. HLA polymorphism in the Havasupai: evidence for balancing selection. *Am. J. Hum. Genet.*, **53**, 943–952.

MARSH, S. G. E., PARHAM, P. AND BARBER, L. D. 2000. *The HLA Facts Book*. London: Academic Press.

MIKKO, S. 1997. *A Comparative Analysis of Genetic Diversity at MHC DRB Loci in Some Ruminant Species*. PhD Thesis, Swedish University of Agricultural Sciences, Uppsala.

MIKKO, S. AND ANDERSSON, L. 1995. Low major histocompatibility complex class II diversity in European and North American moose. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 4259–4263.

MIKKO, S., SPENCER, M., MORRIS, B., STABILE, S. *ET AL*. 1997. A comparative analysis of MHC *DRB3* polymorphism in the American bison (*Bison bison*). *J. Hered.*, **88**, 499–503.

MIKKO, S., ROED, K., SCHMUTZ, S. AND ANDERSSON, L. 1999. Monomorphism and polymorphism at MHC *DRB* loci in domestic and wild ruminants. *Immunol. Rev.*, **167**, 169–178.

NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M. AND GOJOBORI, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.

NEI, M. AND KUMAR, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

O'BRIEN, S. J. AND EVERMANN, J. F. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Ecol. Evol.*, **3**, 254–259.

ORITA, M., IWAHANA, H., KANAZAWA, H., HAYASHI, K. *ET AL*. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 2766–2770.

PAGE, R. D. M. AND CHARLESTON, M. A. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol. Phy. Evol.*, **7**, 231–240.

PATERSON, S. 1998. Evidence of balancing selection at the MHC in a free-living ruminant. *J. Hered.*, **89**, 289–294.

PATERSON, S., WILSON, K. AND PEMBERTON, J. M. 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries*). *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 3714–3719.

RAYMOND, M. AND ROUSSET, F. 1997. GENEPOP (version 3.1d): *Population Genetics Software for Exact Tests and Ecumenicism*. Montpellier, France.

SATTA, Y., O'HUIGIN, C., TAKAHATA, N. AND KLEIN, J. 1994. Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7184–7188.

SCHWAIGER, F. W., WEYERS, E., BUITKAMP, J., EDE, A. J. *ET AL*. 1994. Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. *Mol. Biol. Evol.*, **11**, 239–249.

SETON, E. T. 1929. *Lives of Game Animals*. Doubleday, Doran, Garden City, New York.

SIGURDARDOTTIR, S., BORSCH, C., GUSTAFSSON, K. AND ANDERSSON, L. 1991. Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. *Anim. Genet.*, **22**, 199–209.

SWARBRICK, P. A., SCHWAIGER, F. W., EPPLEN, J. T., BUCHAN, G. S. *ET AL*. 1995. Cloning and sequencing of expressed DRB genes of the red deer (*Cervus elaphus*) MHC. *Immunogenetics*, **42**, 1–9.

THURZ, M. R., THOMAS, H. C., GREENWOOD, B. M. AND HILL, A. V. S. 1997. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genet.*, **17**, 11–12.

VAN DEN BUSSCHE, R. A., HOOFER, S. R. AND LOCHMILLER, R. L. 1999. Characterization of *Mhc-DRB* allelic diversity in white-tailed deer (*Odocoileus virginianus*) provides insight into *Mhc-DRB* allelic evolution within Cervidae. *Immunogenetics*, **49**, 429–437.

Appendix

Frequencies for the 21 MHC class II *DRB* alleles in 12 populations of bighorn sheep, where N is the sample size

Population	N	Ovca-DRB allele																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Mt. Davis	15	—	0.367	—	—	—	0.033	0.200	—	—	0.033	—	0.033	—	—	—	0.333	—	—	—	—	—
Lost Cabin	15	—	0.300	—	—	—	—	0.333	—	—	0.133	—	—	—	—	—	0.233	—	—	—	—	—
Mt Nutt	27	—	0.222	—	—	—	—	0.241	—	0.315	0.019	—	—	—	—	—	0.204	—	—	—	—	—
Kofa Mtns	4	—	—	0.125	—	—	0.125	0.125	0.125	0.250	—	—	—	—	—	—	0.125	0.125	—	—	—	—
Stewart Mtn	14	—	—	—	—	—	0.107	0.036	—	0.107	0.143	—	0.250	0.071	—	—	0.071	0.214	—	—	—	—
Castle Dome	20	—	—	—	—	—	0.100	0.025	—	0.100	0.200	—	0.175	—	—	—	0.200	0.200	—	—	—	—
Old Dad	23	—	0.196	—	—	0.413	—	0.087	—	—	—	0.217	—	—	0.087	—	—	—	—	—	—	—
Eagle Mtns	21	0.071	—	0.143	—	0.095	—	0.190	—	0.238	—	0.071	—	—	—	0.024	0.143	0.024	—	—	—	—
S. Gorgonio	20	—	—	—	—	—	—	0.400	—	—	—	0.100	—	—	—	—	—	0.500	—	—	—	—
San Ysidro	22	—	—	0.068	—	0.704	0.023	—	—	0.136	0.023	0.045	—	—	—	—	—	—	—	—	—	—
Red Rock	25	—	—	—	—	—	0.220	—	—	—	—	—	—	—	—	—	—	—	—	0.360	0.420	—
Wheeler Pk	7	—	—	—	0.143	—	—	—	—	—	—	—	—	—	—	—	—	—	0.571	—	—	0.286
Mean	213	0.007	0.096	0.023	0.005	0.127	0.049	0.141	0.002	0.099	0.045	0.044	0.035	0.005	0.009	0.002	0.106	0.085	0.019	0.042	0.049	0.009