

## ORIGINAL ARTICLE

## Selection on MHC-linked microsatellite loci in sheep populations

F Santucci<sup>1</sup>, KM Ibrahim<sup>2</sup>, A Bruzzone<sup>1,3</sup> and GM Hewit<sup>1</sup><sup>1</sup>School of Biological Sciences, University of East Anglia, Norwich, UK and <sup>2</sup>Department of Zoology, Southern Illinois University, Carbondale, IL, USA

Microsatellites within the major histocompatibility complex (MHC) region have received increasing attention as proxy measures of the level of polymorphism at the *Mhc* genes themselves. We assessed the diversity of microsatellite loci within or in close proximity of the *Mhc* genes in several breeds of domestic sheep (*Ovis aries*) and the wild Mouflon (*Ovis orientalis musimon*). This was compared to variation at other microsatellite loci scattered throughout the sheep genome. Significantly higher number of alleles were observed at the MHC microsatellites. The sheep breeds studied fell into high- and low-diversity group. This grouping is not related to the agricultural use of the breeds, whether for milk, meat or wool. It is, however, correlated with the geographic

origins of the breeds. Southern breeds are genetically more diverse than northern breeds. The observed heterozygosity was in most cases lower than Hardy–Weinberg expectations. The potential impact of selective breeding by man on this is discussed. Neutrality tests indicated that for most of the breeds, the distribution of alleles at the MHC-linked microsatellites are more even than would be expected if the genes were neutral and sampled from populations under drift–mutation equilibrium. Hitchhiking due to tight linkage with alleles at the MHC loci that are under balancing selection is proposed as a possible explanation for this pattern.

*Heredity* (2007) **99**, 340–348; doi:10.1038/sj.hdy.6801006; published online 23 May 2007

**Keywords:** MHC; microsatellites; selection; sheep; neutrality test

## Introduction

High allelic diversity at the major histocompatibility complex (MHC) loci has been reported in many vertebrate species and has been attributed to the vital role MHC proteins play in initiating immune response to parasites and pathogens (Gaudieri *et al.*, 2000). This group of closely linked genes code for proteins that recognize peptides of self or non-self origin and trigger T-cell-mediated specific immune responses. The maintenance of elevated polymorphism at MHC loci is thought to be the consequences of balancing selection of varying nature (Garrigan and Hedrick, 2003). It can take the form of higher fitness of the heterozygote state resulting in homozygote deficit relative to the expectations from neutral theory (Hedrick and Thomson, 1983) and homozygote deficit relative to Hardy–Weinberg proportions (Markow *et al.*, 1993; Black and Hedrick, 1997). There is also evidence that rare-allele-advantage combined with temporally and spatially varying selection, and a modification of this model in which a given allele can confer resistance in either homozygous or heterozygous state (Hedrick *et al.*, 1987) could generate and maintain high levels of MHC diversity (Takahata and Nei, 1990; Slade and McCallum, 1992; Hedrick, 1994; Hill *et al.*, 1994). Recent reviews of the topic list

other explanations such as MHC-specific mate choice (Edwards and Hedrick, 1998; Jordan and Bruford, 1998; Garrigan and Hedrick, 2003).

Microsatellite loci located within the MHC region have received increasing attention as proxies for inferring the level of polymorphism at MHC loci (Ammer *et al.*, 1992; Ellegren *et al.*, 1993; Schwaiger *et al.*, 1993, 1994). The analysis of variation at these loci has produced contrasting results. High polymorphism and heterozygote excess relative to neutral expectations were observed in cattle (van Haeringen *et al.*, 1999) and in free living Soay sheep (Paterson, 1998), while no deviation from Hardy–Weinberg equilibrium was found in a sample of Dutch people (Moghaddam *et al.*, 1998). Low levels of variation and no evidence of selection were reported in bighorn sheep (Boyce *et al.*, 1997).

Although the evidence for balancing selection acting on MHC genes is considerable, either from its impact on diversity at the MHC loci themselves or on MHC-linked microsatellites, a common problem in the statistical tests to establish this has been the confounding effect of demographic events. A commonly used method is the Ewens–Watterson test (Ewens, 1972; Watterson, 1978), which compares observed allelic profiles with predictions of the neutral theory assuming the population being tested has maintained a constant size. The sheep breeds we studied will have undergone bottlenecks as a result of active breeding. This can have a confounding effect on the Ewens–Watterson test. Screening other neutral markers in addition to the MHC or the proxy-MHC loci can overcome this effect (Schlotterer *et al.*, 2004). Non-MHC microsatellite loci make good candidates for this purpose. The impact of selection can be

Correspondence: Current address. Dr F Santucci, Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Microbiologia, Università degli Studi di Perugia, Via del Giochetto, 06122 Perugia, Italy.

E-mail: fiammetta@fastmail.fm

<sup>3</sup>Dipartimento SAVA, Università del Molise, Campobasso 86100, Italy.

Received 6 July 2006; revised 4 April 2007; accepted 7 April 2007; published online 23 May 2007

distinguished by comparing the results of the test for the neutral markers and for the MHC loci. Selection typically affects specific loci (Garrigan and Hedrick, 2003), while the effect of bottlenecks and other demographic events on diversity is likely to be genome wide.

We have investigated polymorphism at microsatellite loci linked to MHC genes (henceforth referred to as MHC microsatellites) in several breeds of domestic sheep and in one of their wild relatives, the European Mouflon, *Ovis orientalis musimon* (Corbet and Hill, 1980). This has been compared to variation at other microsatellite loci distributed across the sheep genome (non-MHC microsatellites). The latter data are from a much broader collaborative screening project between us and other researchers on the genetic diversity of European sheep breeds, whose detailed reports are forthcoming. The history of the breeds we studied encompasses varying durations and intensities of artificial selection. This has enabled us to address specific questions about the impact of managed breeding on the genetic diversity of MHC microsatellites relative to non-MHC microsatellites.

## Materials and methods

The following sheep breeds were studied: Chios from Greece, Comisana from Italy, Friesland from The Netherlands, Merino from Spain, Sarda from Italy, Skudde from Germany and Soay from Great Britain. Several samples from the wild sheep known as Mouflon (*O. o. musimon*) in Sardinia and the Caroux-Espinouse Massif in southern France were also included. In Europe, Mouflons are to be found in the wild only in the islands of Sardinia and Corsica. However, descendants of these wild populations have been introduced to isolated mountain regions in the Alps and elsewhere for hunting purposes (Ryder, 1984). Friesland, Chios and Sarda are milk breeds. Comisana and Skudde are meat breeds, while Merino is a wool breed. Soay is a free-living primitive breed. The number of sheep that were genotyped is shown in Table 2. Between one and three flocks were sampled for each breed.

DNA was isolated from white blood cells or tissue by classical phenol/chloroform extraction methods (Sambrook *et al.*, 1989) and from hair using a chelex protocol (Zhang and Hewitt, 1998). The genotypic makeup of the sheep at three microsatellites loci that are closely linked to MHC genes was determined. These are: (1) OLADRB (Schwaiger *et al.*, 1993) located in the second intron of *OvarDRB1*, a class II MHC gene; (2) OLADRBps located in the intron between the fifth and the sixth exons of *OLADRBps* (Blattman and Beh, 1992); and (3) OMHC1 in the third intron of the *OMHC1* class I locus (Groth and Wetherall, 1994). OLADRB is an interrupted and complex microsatellite with (gt)<sub>n</sub>(ga)<sub>m</sub> as the basic motif. OMHC1 has a (ca) motif. OLADRBps has an uninterrupted (ac) motif (Table 1).

Polymerase chain reaction amplifications were carried out in a 9700 PE thermal cycler. Amplification was in a 10  $\mu$ l reaction volume containing 2.3 mM MgCl<sub>2</sub>, 250  $\mu$ M dNTPs, 1.8  $\times$  AmpliTaq Gold Buffer, 0.18  $\mu$ M of each primer and 2.5 U of AmpliTaq Gold (Perkin Elmer). The thermal profile had an initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturing at 94°C for 1 min, annealing at 60°C for 1 min (OMHC1 and OLADRB), or 55°C for 30 s (OLADRBps), and extension

at 72°C for 1 min. The final stage comprised an extension reaction for 10 min at 72°C.

Summary data pertaining to an additional 12 non-MHC microsatellite loci have been included in this study for the purpose of comparing the allelic and genotypic distributions in MHC and non-MHC microsatellite loci. The loci are: OarFCB48, MAF214, ETH225, ILSTS005, ILSTS011, ILSTS28, SRCRSP1, SRCRSP3, SRCRSP5, SRCRSP7, SRCRSP8, SRCRSP9 as characterized in (Steffen *et al.*, 1993; Arevalo *et al.*, 1994; Bhebe *et al.*, 1994; Bishop *et al.*, 1994; Crawford *et al.*, 1995; Kemp *et al.*, 1995; Brezinsky *et al.*, 1993a,b). Conditions used for the amplification and genotyping of the non-MHC microsatellites were similar to those for the MHC microsatellite loci. Allele size scoring for both sets of data was carried out in an ABI 377 automated sequencer, using the ABI GeneScan and Genotyper software. An Ovine DNA standard was included in every gel. Allele frequencies and number of alleles ( $N_a$ ) were calculated using Genepop ver.3.1d (Raymond and Rousset, 1995) as were the expected ( $H_e$ ) and observed heterozygosities ( $H_o$ ). Allelic diversity of the MHC and non-MHC microsatellite loci was compared using allelic richness, the number of alleles per locus. Measures of allelic richness were corrected for variation in sample size by using the rarefaction method as implemented in FSTAT (Goudet, 2000). This essentially compares the number of alleles per locus per breed estimated from  $n$  diploid genotypes as in El Mousadik and Petit (1996), and Hurlbert (1971), where  $n$  is the smallest number of sheep genotyped in any of the breeds. There were 14 OLDRB genotypes available for Skudde and hence all the allelic richness estimates obtained for both the MHC and non-MHC microsatellite loci are per 14 diploid individuals. Slatkin (1994) implementation of the Ewens-Watterson neutrality test (Ewens, 1972; Watterson, 1978) was performed using PyPop (Lancaster *et al.*, 2003).

Ewens (1972) derived a formula for the expected number of neutral alleles at equilibrium between mutation and loss of new alleles due to drift. In a sample of  $2N$  genes, this is given by:  $\sum_{i=1}^{2N} \theta / (\theta + i)$ , where  $\theta = 4N_e\mu$ . Thus, if  $\theta$ , the product of mutation rate and effective population size is small, the expected number of different alleles approaches 1 and if  $\theta$  is very large, it approaches  $2N$  (Hedrick, 2000). Watterson (1978) developed a test for neutrality based on this. The test calculates the observed Hardy-Weinberg homozygosity ( $F$ ) in a sample of  $2N$  genes containing  $k$  different alleles as  $\sum p_i^2$ , where  $p_i$  is the frequency of allele  $i$ . This is then compared with the expected homozygosity,  $\hat{F}$ , in a similar sample of  $2N$  genes also containing  $k$  alleles drawn per Ewens' expectations assuming neutrality and drift-mutation equilibrium. A significant difference between the two is taken as evidence for selection. If the observed homozygosity is lower, indicating the frequencies of the  $k$  alleles in the sample is even more than expected under neutrality, balancing selection can be inferred. Deviation in the opposite direction may be attributed to purifying selection (Hedrick, 2000).

## Results

Table 1 shows information on the microsatellites typed, reporting for each locus the motif, size range, number of alleles and allelic richness. Both the MHC and the non-

**Table 1** Locus information

Locus	Origin	NCBI	Motif	Size range (bp)	Alleles	Allelic richness
<i>MHC microsatellite loci</i>						
OLADRB	Ovine	U00204	(gt) <sup>20</sup> (ga) <sup>15</sup> (gc) <sup>5</sup> (ga) <sup>4</sup> (gc) <sup>4</sup> n(ag) <sup>4</sup> n(ag) <sup>4</sup> n(ag) <sup>4</sup> n(ag) <sup>5</sup>	190–316	44	9.62
OLADRBps	Ovine	AH003856	(ca) <sup>16</sup>	265–299	18	7.42
OMHC1	Ovine	U21953	(gc) <sup>3</sup> n(ca) <sup>24</sup>	168–212	14	7.18
				Mean (s.e.)		8.07 (1.34)
<i>Non-MHC microsatellite loci</i>						
OarFCB48	Ovine	M82875	(tg) <sup>11</sup> n(tg) <sup>3</sup>	145–171	13	5.03
MAF214	Ovine	M88160	(gt) <sup>10</sup> n(gt) <sup>7</sup> n(gt) <sup>4</sup> n(gt) <sup>4</sup> n(tg) <sup>5</sup> n(tg) <sup>8</sup>	184–270	17	5.55
ETH225	Bovine	Z14043	(tg) <sup>4</sup> n(ca) <sup>18</sup> n(ct) <sup>3</sup>	135–153	10	5.69
ILSTS005	Bovine	L23481	(tg) <sup>4</sup> n(at) <sup>6</sup> (gt) <sup>9</sup> n(ta) <sup>7</sup>	180–204	13	3.54
ILSTS011	Bovine	L23485	(tc) <sup>7</sup> n(ca) <sup>11</sup>	264–290	14	4.79
ILSTS28	Bovine	L37211	(catt) <sup>3</sup> (ca) <sup>7</sup>	125–171	18	5.45
SRCRSP1	Caprine	L22192	(ac) <sup>24</sup>	116–148	15	3.75
SRCRSP3	Caprine	L22195	(tg) <sup>7</sup> (ta) <sup>2</sup> (tg) <sup>11</sup>	115–197	18	6.17
SRCRSP5	Caprine	L22197	(at) <sup>3</sup> (gt) <sup>16</sup> n(tc) <sup>3</sup>	139–151	7	6.10
SRCRSP7	Caprine	L22199	(ct) <sup>3</sup> n(gt) <sup>7</sup> (at) <sup>9</sup> n(at) <sup>4</sup>	158–214	23	4.55
SRCRSP8	Caprine	L22193	(aac) <sup>3</sup> n(at) <sup>6</sup> (gt) <sup>9</sup> n(tg) <sup>5</sup> n(tg) <sup>6</sup> (at) <sup>4</sup> (tg) <sup>3</sup> (at) <sup>14</sup> n(tga) <sup>3</sup>	213–247	18	4.32
SRCRSP9	Caprine	L22000	(ag) <sup>8</sup> n(tg) <sup>20</sup> n(tg) <sup>3</sup>	109–135	12	5.44
				Mean (s.e.)		5.03 (0.86)

For each microsatellite locus, the species for which the primers were originally developed, the GeneBank accession number of the microsatellite sequences (NCBI), the repeat motif and size range are shown. Gross number of alleles per locus and number of alleles corrected for sample size (allelic richness) are also shown.

**Table 2** Summary statistics of allelic and genotypic distributions of three MHC-embedded microsatellite loci and 12 non-MHC microsatellite loci in European sheep breeds

Breed locus	Sarda	Comisana	Merino	Chios	Mouflon	Skudde	Soay	Friesland
OLADRB	(31)	(25)	(28)	(27)	(28)	(14)	(27)	(30)
N <sub>a</sub>	14	20	14	17	11	7	8	7
H <sub>o</sub>	0.23	0.76	0.52	0.52	0.43	0.71	0.96	0.67
H <sub>e</sub>	0.87	0.93	0.86	0.71	0.77	0.81	0.84	0.78
P	<0.01	<0.01	<0.01	<0.01	<0.01	0.121	0.681	<0.01
OLADRBps	(40)	(28)	(37)	(32)	(25)	(22)	(26)	(22)
N <sub>a</sub>	15	13	7	9	7	7	6	6
H <sub>o</sub>	0.88	0.79	0.89	0.56	0.44	0.73	0.85	0.73
H <sub>e</sub>	0.91	0.83	0.83	0.72	0.64	0.72	0.80	0.77
P	0.058	0.065	0.026	<0.01	<0.01	0.014	0.120	0.706
OMHC1	(40)	(31)	(39)	(31)	(26)	(31)	(22)	(24)
N <sub>a</sub>	13	12	10	10	6	7	8	5
H <sub>o</sub>	0.78	0.71	0.79	0.74	0.38	0.68	0.59	0.46
H <sub>e</sub>	0.86	0.82	0.83	0.82	0.64	0.71	0.71	0.69
P	0.028	<0.01	0.021	0.012	<0.01	0.069	0.273	0.138
<i>Non-MHC</i>								
<sup>1</sup> N <sub>a</sub>	7.00	7.66	6.55	6.00	6.50	5.80	4.36	4.63
<sup>2</sup> H <sub>o</sub>	0.57 (0.18)	0.63 (0.16)	0.54 (0.16)	0.53 (0.20)	0.51 (0.23)	0.62 (0.16)	0.43 (0.15)	0.47 (0.24)
<sup>2</sup> H <sub>e</sub>	0.64 (0.13)	0.68 (0.13)	0.63 (0.08)	0.61 (0.19)	0.68 (0.16)	0.67 (0.08)	0.45 (0.17)	0.51 (0.21)
<sup>3</sup> n <sub>P</sub>	3	1	2	3	9	3	0	1

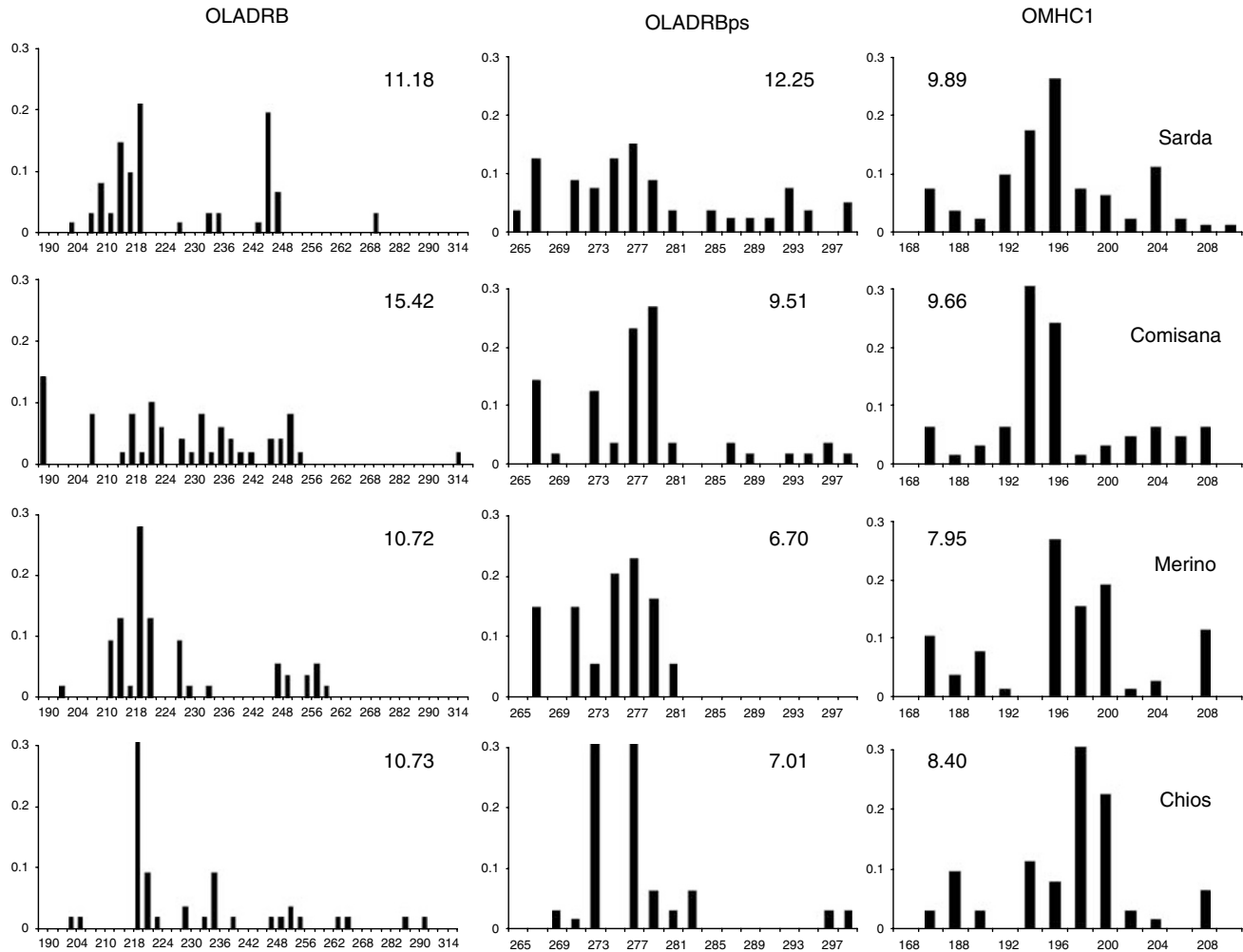
In brackets are number of individuals genotyped at each locus. N<sub>a</sub>, H<sub>o</sub> and H<sub>e</sub> are number of alleles, observed heterozygosity and expected heterozygosity, respectively. P is the probability from test of conformity to Hardy–Weinberg expectations. In the case of the 12 non-MHC loci, (1) the mean number of alleles per locus, (2) the mean and the standard deviation of observed and expected heterozygosities and (3) the number of loci that showed statistically significant deviation from Hardy–Weinberg proportions are given.

MHC group contain simple and complex repeats, some perfect and others imperfect.

#### Allelic diversity

Summary statistics of allelic and genotypic distributions of the 3 MHC-embedded microsatellite loci and the 12

non-MHC microsatellite loci are shown in Tables 1 and 2. Focussing on the MHC microsatellites, the average number of alleles per breed, without correcting for differences in sample sizes, is 12.25±4.83 at OLADRB, 8.75±3.41 at OLADRBps and 8.88±2.85 at OMHC1. The total number of alleles across all breeds is 44 for OLADRB, 18 at OLADRBps and 14 at OMHC1; OLADRB



**Figure 1** Histograms showing allele frequencies at the OLADRB, OLADRBps and OMHC1 microsatellites in the high-diversity group (Sarda, Comisana, Merino and Chios). Allelic richness estimates (see text) are also shown.

is the most polymorphic locus, while OLADRBps and OMHC1 show comparable allelic diversities.

The average number of alleles per locus in each breed ranges between 15 in Comisana and 6 in Friesland. Sarda, Comisana, Merino and Chios have consistently higher number of alleles in any of the three loci compared to Mouflon, Skudde, Soay and Friesland. This pattern is also obvious in Figures 1 and 2, where the frequency distributions of MHC-microsatellite alleles in the eight sheep breeds are shown as histograms. Thus, based on allelic diversity at the MHC-microsatellite loci, the eight sheep breeds can be grouped into two. A high-diversity group comprising Sarda, Comisana, Merino and Chios with an average of  $38.5 \pm 6.24$  alleles per breed in the three loci combined and a less variable second group comprising Mouflon, Soay, Skudde and Friesland with  $21.25 \pm 2.25$  alleles (Table 2). This grouping also holds for each of the three loci individually. The average number of OLADRB alleles in the first group is  $16 \pm 2.87$  compared to  $8.25 \pm 1.89$  in the second group. Similarly, the comparisons are  $11 \pm 3.65$  versus  $6.5 \pm 0.58$  for OLADRBps and  $11.25 \pm 1.5$  versus  $6.5 \pm 1.29$  for OMHC1.

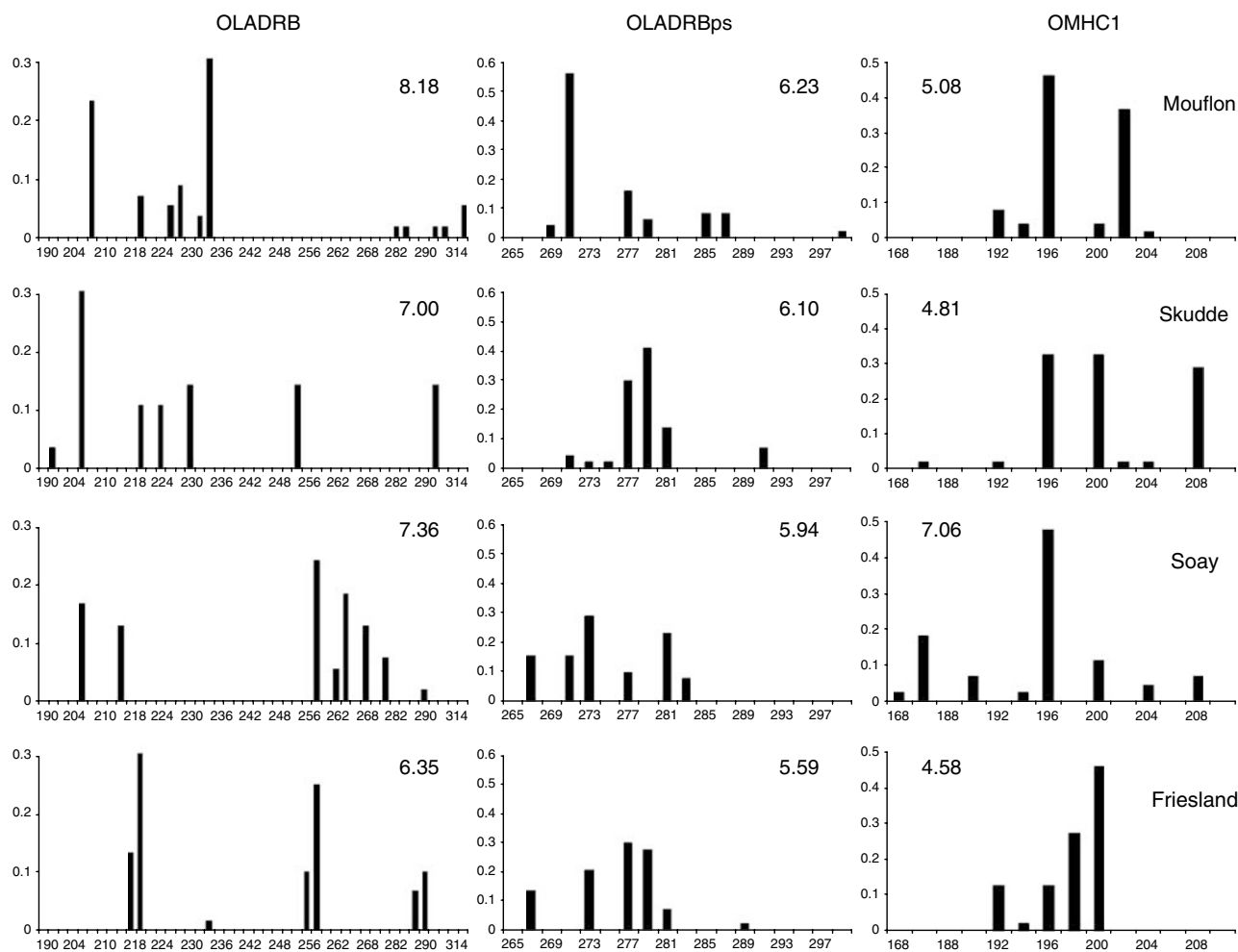
The above trend is confirmed when allelic richness, corrected for differences in sample sizes, is used as a

measure of diversity (Table 1, Figures 1 and 2). Allelic richness of the eight breeds averaged across the three MHC-microsatellite loci is  $8.07 \pm 1.34$ . The high- and low-diversity breeds had average allelic richness of 9.95 and 6.19, respectively. The difference between the two was found to be statistically significant ( $P = 0.014$ ) in a bootstrap test of 1000 replicates.

The average number of alleles per locus at the 12 non-MHC microsatellite loci is consistently lower than the number of alleles at any one of the three MHC microsatellites in all the breeds but Mouflon. In Mouflon, six non-MHC loci had number of alleles ranging between seven and nine compared to the seven and six alleles at OLADRBps and OMHC1, respectively. However, averaged across loci, this breed too is more diverse at the MHC-microsatellite loci than the non-MHC loci. Allelic richness of the non-MHC loci, at  $5.03 \pm 0.86$  alleles per locus, is also significantly lower than in the MHC-microsatellite loci (Table 1).

#### Heterozygosity

Table 2 shows the frequency of  $H_o$  at each of the three MHC-microsatellite loci for each breed. Averaged across



**Figure 2** Histograms showing allele frequencies at the OLADR, OLADRBps and OMHC1 microsatellites in the low-diversity group (Mouflon, Skudde, Soay, Friesland). Allelic richness estimates (see text) are also shown.

breeds ( $0.60 \pm 0.22$  at OLADR,  $0.73 \pm 0.16$  at OLADRBps and  $0.64 \pm 0.15$  at OMHC1), it is consistently lower than the average expected heterozygosity ( $0.82 \pm 0.07$  at OLADR,  $0.78 \pm 0.08$  at OLADRBps,  $0.76 \pm 0.08$  at OMHC1). At individual loci in each breed too, this is predominantly the case. This is the expected trend for the sampling scheme of multiple, separately managed flocks per breed that we adopted to be able to assess breed-wide levels of genetic diversity in European sheep.

When the genotypic distributions at the MHC-microsatellite loci were tested for conformity to Hardy-Weinberg proportions, all the breeds except Soay showed significant deficits of heterozygotes in at least one of the three loci ( $P < 0.05$ , Table 2). This is particularly evident at OLADR. All the breeds except Soay and Skudde showed highly significant ( $P < 0.01$ ) heterozygote deficit at this locus. In the semiwild Soay, where our sample is effectively from a single flock, there is no deviation from the Hardy-Weinberg proportions at any of the three loci. The opposite is true in the case of the wild species Mouflon. All three loci showed highly significant heterozygote deficit. This sample included sheep belonging to populations from two localities in Sardinia and France. When tested separately, despite the smaller

sample sizes, the deviations from Hardy-Weinberg proportions were found to be non significant at all three loci confirming that the homozygote excess in the pooled data was due to the Wahlund effect.

In the 12 non-MHC microsatellite loci, the average  $H_o$  across loci and breed was 0.54 (range 0.43–0.71); the range per breed was 0.43–0.63. Hardy-Weinberg tests showed significant heterozygote deficit in at least one out of the 12 non-MHC microsatellite loci in all the breeds with the exception of Soay. All 12 loci were at equilibrium in Soay. In Mouflon, nine non-MHC microsatellite loci revealed significant heterozygote deficit, confirming the Wahlund effect observed in the MHC microsatellites. The average expected and observed heterozygosities are higher in the MHC microsatellites as a group compared to the non-MHC loci primarily due to the allelic richness of the former. This is not the case for Mouflon. Several factors such as breeding practices, flock structure and breed history that may have affected the above allelic and genotypic diversity patterns are discussed below.

Table 3 shows results of the homozygosity tests of neutrality for the three MHC-microsatellite loci in the eight breeds. Till now, as the predominant trend in the

**Table 3** Observed Hardy–Weinberg homozygosity ( $F$ ) was calculated as  $\sum p_i^2$ , where  $p_i$  is the frequency of allele  $i$

Breed	Locus	$N_a$	$F$	$\hat{F}$	$F$ deviate	$P$
Sarda	OLADRB	14	0.1228	0.1499	-0.6238	0.2874
	OLADRBps	15	0.0903	0.1643	-1.4494	0.0028*
	OMHC1	13	0.1409	0.1937	-0.8551	0.1657
	Non-MHC loci	6		6		0
Comisana	OLADRB	20	0.0720	0.0927	-1.0187	0.0875
	OLADRBps	13	0.1684	0.1718	-0.0687	0.5804
	OMHC1	12	0.1764	0.1955	-0.3136	0.4658
	Non-MHC loci	4		8		0
Merino	OLADRB	14	0.1390	0.1568	-0.3939	0.4231
	OLADRBps	7	0.1702	0.3568	-1.5462	0.0023*
	OMHC1	10	0.1654	0.2559	-1.0530	0.0872
	Non-MHC loci	6		6		0
Chios	OLADRB	17	0.2929	0.1204	5.5528	0.9985*
	OLADRBps	9	0.2817	0.2698	0.1351	0.6516
	OMHC1	10	0.1811	0.2400	-0.7596	0.2231
	Non-MHC loci	3		9		1
Mouflon	OLADRB	11	0.2430	0.2091	0.5265	0.7728
	OLADRBps	7	0.3576	0.3267	0.2928	0.6936
	OMHC1	6	0.3558	0.3789	-0.1919	0.5179
	Non-MHC loci	3		9		0
Skudde	OLADRB	7	0.1888	0.2778	-1.1265	0.0675
	OLADRBps	7	0.2810	0.3165	-0.3540	0.4468
	OMHC1	7	0.2934	0.3434	-0.4390	0.4055
	Non-MHC loci	2		10		0
Soay	OLADRB	8	0.1626	0.2912	-1.3758	0.0105*
	OLADRBps	6	0.1990	0.3789	-1.4903	0.0048*
	OMHC1	8	0.2862	0.2756	0.1287	0.6496
	Non-MHC loci	7		7		1
Friesland	OLADRB	7	0.2161	0.3409	-1.1064	0.0799
	OLADRBps	6	0.2273	0.3655	-1.2133	0.0507
	OMHC1	5	0.3151	0.4365	-0.8916	0.1862
	Non-MHC loci	4		8		1

The expected homozygosity, ( $\hat{F}$ ) in a sample of the same size and containing the same number of different alleles is calculated per Ewens' expectations using the package PyPop (Lancaster *et al.*, 2003). The  $P$ -values represent the statistical significance of the difference between the two homozygosity estimates. Asterisks show values significant at 0.05 level. Similar analyses were carried out for the 12 non-MHC microsatellite loci. The number of loci where  $F$  is higher than  $\hat{F}$  is shown in the  $F$  column, while the opposite is given in the  $\hat{F}$  column. The number of loci where the difference between the two was statistically significant is shown in the  $P$  column.

tests has revealed few statistically significant differences between the observed Hardy–Weinberg homozygosities and  $\hat{F}$ , these results can be interpreted as evidence for neutrality of the MHC-microsatellite loci. However, there is a clear trend of lower Hardy–Weinberg homozygosity compared to neutral expectations. In 19 out of the 24 tests, observed  $F$  is lower than  $\hat{F}$  indicative of allelic distributions more even than neutral expectation.

Among the four tests where the deficit in  $F$  compared to  $\hat{F}$  is statistically significant, two involve Soay. Thus, in the only breed where the Hardy–Weinberg tests confirmed random mating, there is significant evidence for selection at the OLADRB and OLADRBps loci but not at OMHC1.

Among the five neutrality tests that revealed higher Hardy–Weinberg homozygosity compared to that expected under neutrality, three involve the wild species Mouflon and the semiwild Soay. The remaining two pertain to Chios (Table 3). As with the tests for deviations from Hardy–Weinberg expectations, the Mouflon results appear to be artefacts of sampling of two separate populations from Sardinia and the French Alps. When the 26 Sardinia Mouflons were tested separately, the observed homozygosity was lower than the expected. Similarly, when 15 Chios sheep from a single flock in Athalsa, Cyprus were tested separately, the observed homozygosity at all three MHC microsatellites was lower than the expected. It is thus apparent that, when the effect of multiple sampling is taken into account, all but one (Soay, OMHC1) of the 24 comparisons of  $F$  and  $\hat{F}$  are indicative of allelic distributions more uniform than expected for neutral loci. However, when tested for statistical significance, only four could be confirmed. Fisher (1932) combined probability test was used on the set of  $P$ -values in the 24 tests and produced a highly significant  $P$ -value ( $X^2 = 94.35$ , d.f. = 48,  $P < 0.001$ ).

Table 3 also summarizes the results of tests of neutrality for the 12 non-MHC microsatellite loci in the eight breeds, a total of 96. Observed  $F$  was lower in 63 tests, but none were significant. In contrast to the major trend at the MHC microsatellites, the only statistically significant deviations, of which there were three, were all in the direction of higher observed homozygosity. This and the fact that over a third of the tests revealed no deviation from neutrality lead to the conclusion that allelic frequencies at these loci do not show a similar pattern to the MHC microsatellite loci.

## Discussion

MHC-embedded microsatellite loci in European sheep breeds were found to be highly polymorphic as is the case in a number of other domestic ungulates (Paterson, 1998; van Haeringen *et al.*, 1999). Both the number of alleles per locus and the genotypic diversity as measured by the expected number of heterozygotes are consistently higher compared to other microsatellite loci dispersed throughout the sheep genome. We infer from this that the loci may have been affected by selection at the neighbouring *Mhc* genes where diversity is also high. Balancing selection has been proposed as the mechanism responsible for the maintenance of high levels of polymorphism at MHC loci. Other mechanisms, such as assortative mating, maternal–fetal interactions and selective abortions have also been proposed (for a review, see Hedrick, 1994; Jordan and Bruford, 1998; Garrigan and Hedrick, 2003). The hitchhiking effect of the tight linkage of the MHC microsatellites to these genes would lead to higher polymorphism too.

There is indication that repeat type and number of repeats have an influence on microsatellite variability (Ellegren, 2000; Dieringer and Schlotterer, 2003). To detect if this is the case in our two sets of data, we

compared all available information on the typed loci. As most of the non-MHC microsatellites used in this study have been derived in cattle and goat mapping projects no information is available on their structure in sheep. This should be born in mind when considering any conclusions drawn from comparison of their structure to the MHC group. Table 1 shows no simple pattern when looking at repeat structure or, type and number of repeats. Both the MHC and non-MHC group contain repeat motifs equally complex, often interrupted and characterized by a high number of repeats. Interestingly, the MHC microsatellite with higher number of alleles also shows a wider range in allele size. This seems to suggest that alleles at this locus follow the pattern observed at MHC genes, where alleles have been retained over large evolutionary periods (Klein, 1987). In support of this view, the basic  $(gt)_n(ga)_m$  motif of OLADRB is preserved in the same location over a large evolutionary span, having been observed in cattle, sheep, goat, reindeer and humans (Schwaiger *et al.*, 1994; Schwaiger and Epplen, 1995).

The eight sheep breeds we studied can be grouped into a high-diversity group and a low-diversity group based on allelic richness at the MHC microsatellite loci. Sarda (milk breed), Comisana (meat), Merino (wool) and Chios (milk) comprise the diverse group. Mouflon (wild), Soay (semiwild), Skudde (meat) and Friesland (milk) are in the group with fewer alleles. Thus, the grouping does not appear to be linked to the purpose for which the breeds were bred and managed. However, this pattern may be associated with the geographic origins of the breeds. The high-diversity group comprises primarily of southern breeds, while the low-diversity group are of northern European origins. This grouping reflects what has been shown by the main population study (Lawson-Handley *et al.*, submitted), where breeds from southern Europe tend to exhibit more variation than northern ones. This has been related to a loss of diversity further away not only from the centre of domestication, but also from southern glacial refugia, from where populations may have recolonized northern areas, as observed in several wild species (Hewitt, 2000).

In all the breeds we studied, there is higher diversity in the MHC microsatellites compared to non-MHC microsatellites, above and beyond that associated with the geographic origin of the breeds. We also report that the observed heterozygosity at sheep MHC microsatellites is lower than that expected under Hardy-Weinberg equilibrium in the majority of the loci and breeds. Three factors will have contributed to this. Because our major aim was to assess the overall genetic diversity of European sheep breeds, we sampled one to three flocks for each breed. Clearly, the Wahlund effect would cause heterozygote deficit in the multiflock samples, where the flocks have been separated long enough. However, heterozygote deficit resulting from the Wahlund effect is expected to show across all loci for each affected breed. The two breeds Mouflon and Soay are worth particular attention in this regard. Both are essentially free ranging wild breeds and both conform to Hardy-Weinberg expectations when the Sardinia and France samples of Mouflon are treated separately. In the other breeds, there is variation between loci within breeds on whether there is heterozygote deficit or not. For example, both Skudde and Merino show higher observed heterozygosity at

OLADRBps and lower heterozygosity at OLADRB and OMHC1. From this, we conclude that substructuring resulting from the multiflock-sampling scheme does not fully account for the heterozygote deficit in all the breeds.

The presence of null alleles is another potential cause of deficit in observed heterozygosity. In the breeds, where heterozygote deficit is apparent, most of the loci show it. This is true in the 12 non-MHC microsatellite loci too, regardless of the repeat structure. In addition, the wild Mouflons, when tested separately for the two sampling locations, and the semiwild Soay showed no heterozygote deficit. These observations lead to the conclusion that null alleles are unlikely to have contributed to the observed genotypic distribution.

A third contributing factor is the common use of a single or few rams in many flocks to sire lambs over several generations. This would result in heterozygote deficit relative to Hardy-Weinberg expectations, either because our samples were not from a single generation cohort or because the sires used tend to be more homozygous relative to a male selected at random from the parental generation. That the only two populations that live in wild conditions (semiwild in the case of the Soay from St Kilda Island) showed the least discrepancy between observed and expected heterozygosities argues that the low observed heterozygosity in the other breeds is most likely the consequence of such breeding practices. In striking contrast to its high observed and expected heterozygosity at MHC microsatellites, the Soay breed showed the lowest heterozygosity values in the population study (Lawson-Handley *et al.*, submitted). The Soay population is known to have suffered population crashes (Bancroft *et al.*, 1995; Charbonnel and Pemberton, 2005). This indicates that even populations undergoing bottlenecks are able to retain their MHC variation, as long as random mating is permitted.

The results of the neutrality tests showed that at all the MHC-microsatellite loci and for most of the breeds, the allele distributions are even more than would be expected if the genes were neutral and sampled from populations under drift-mutation equilibrium. However, when tested for statistical significance, this trend is confirmed only in the case of Soay. It is nonetheless noteworthy that the evidence for selection is apparent in Soay, a nonmanaged breed, which does not show Hardy-Weinberg discrepancies. This again seems to support the fact that the lack of statistically significant *F*-values at MHC microsatellites in the other breeds could be due to the confounding effect of managed breeding, which could reduce heterozygosity while maintaining within breed diversity. Thus, it is apparent that the predominant trend in the test of neutrality of these loci is indicative of allelic distributions more uniform than expected of neutral loci in drift-mutation equilibrium, although statistical significance is attained only in the case of Soay. The Ewens-Watterson test has relatively low statistical power, as is the case with other tests of conformity to population genetic models including the Hardy-Weinberg equilibrium (Hedrick, 2000). The consistent trend of lower observed heterozygosity compared to the Ewens-Watterson expectation could reasonably be attributed to the effect of selection. Indeed, Worley *et al.* (2006) have shown, as in our case, that evidence for selection becomes more widely significant when a global

statistics combining multilocus signal is considered as opposed to single loci. While the combining of probabilities following Fisher's method does not exactly address this issue, the fact that a highly significant *P*-value was obtained is indicative of lack of resolution in the tests statistics used for individual loci. It is also noteworthy that the Ewens–Watterson test of neutrality is a composite test of the combined effect of the observed number of alleles, whether higher or lower than expected under the infinite allele mutation model and, their frequency distribution, whether more even or skewed compared to expectations under neutrality. With respect to the former, Schlotterer *et al.* (2004) have demonstrated using simulations that, for a broad range of parameter values, the theoretical expectation of the number of alleles from the infinite allele model tends to be an overestimate. Despite this, the Ewens–Watterson tests consistently yielded deviations in the direction of excess compared to theoretical expectations. This further strengthens our inference that the observed number of alleles and their frequency distributions have been affected by selection.

Hitchhiking due to tight linkage with alleles at the MHC loci that are under balancing selection is proposed as the most likely explanation for this pattern of deviation from neutrality and at least for OLADRB there is some evidence for it (Schwaiger and Epplen, 1995). Previous studies have also identified balancing selection as the most likely form of selection to cause deviations from neutrality in the direction observed bearing in mind the role of the MHC loci to which these microsatellites are tightly linked (Paterson, 1998). It should also be noticed that OLADRBps, a microsatellite located in the fifth intron of a pseudogene of the *OvarDRB1*, shows less variability than OLADRB, located in the second intron of the functional *OvarDRB1*, again supporting the hitchhiking effect on the level of variation. In conclusion, while these data show evidence of balancing selection at the MHC microsatellite, they also highlight the confounding effect of managed breeding on this pattern and the need for more sophisticated tests to unravel its causes.

## Acknowledgements

We are grateful to A Alberti, C Ehling, MJ Ferrano, E Lenstra, J Pemberton, E Petit, F Pilla, A Valentini, who have generously contributed samples to this project. We also thank M Bruford and L Chicki for critical reading and helpful suggestions, and S Paterson for primer information. This work was supported by the European Community, under Framework IV, grant BIO4 CT 961189.

## References

Ammer H, Schwaiger FW, Kammerbauer C, Gomolka M, Arriens A, Lazary S *et al.* (1992). Exonic polymorphism vs intronic simple repeat hypervariability in Mhc-Drb genes. *Immunogenetics* **35**: 332–340.

Arevalo E, Holder DA, Derr JN, Bhebhe E, Linn RA, Ruvuna F *et al.* (1994). Caprine microsatellite dinucleotide repeat polymorphisms at the Sr-Crsp-1, Sr-Crsp-2, Sr-Crsp-3, Sr-Crsp-4 and Sr-Crsp-5 loci. *Animal Genetics* **25**: 202.

Bancroft DR, Pemberton JM, Albon SD, Robertson A, Maccoll ADC, Smith JA *et al.* (1995). Molecular-genetic variation and individual survival during population crashes of an un-

managed ungulate population. *Philos Trans R Soc Lond B Biol Sci* **347**: 263–273.

Bhebhe E, Kogi J, Holder DA, Arevalo E, Derr JN, Linn RA *et al.* (1994). Caprine microsatellite dinucleotide repeat polymorphisms at the Sr-Crsp-6, Sr-Crsp-7, Sr-Crsp-8, Sr-Crsp-9 and Sr-Crsp-10 loci. *Anim Genet* **25**: 203.

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA *et al.* (1994). A genetic-linkage map for cattle. *Genetics* **136**: 619–639.

Black FL, Hedrick PW (1997). Strong balancing selection at HLA loci: evidence from segregation in South Amerindian families. *Proc Nat Acad Sci USA* **94**: 12452–12456.

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

Boyce WM, Hedrick PW, MuggliCockett NE, Kalinowski S, Penedo MCT, Ramey RR (1997). Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics* **145**: 421–433.

Brezinsky L, Kemp SJ, Teale AJ (1993a). 5 Polymorphic bovine microsatellites (Ilsts010-014). *Anim Genet* **24**: 75–76.

Brezinsky L, Kemp SJ, Teale AJ (1993b). Ilsts005 – a polymorphic bovine microsatellite. *Anim Genet* **24**: 73.

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

Corbet GB, Hill JE (1980). *A World List of Mammalian Species*. British Museum (Natural History): London.

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG *et al.* (1995). An autosomal genetic-linkage map of the sheep genome. *Genetics* **140**: 703–724.

Dieringer D, Schlotterer C (2003). Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* **13**: 2242–2251.

Edwards SV, Hedrick PW (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol Evol* **13**: 305–311.

El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L) Skeels] endemic to Morocco. *Theor Appl Genet* **92**: 832–839.

Ellegren H (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* **16**: 551–558.

Ellegren H, Davies CJ, Andersson L (1993). Strong association between polymorphisms in an intronic microsatellite and in the coding sequence of the Bola-Drb3 gene – implications for microsatellite stability and Pcr-based Drb3 typing. *Anim Genet* **24**: 269–275.

Ewens WJ (1972). Sampling theory of selectively neutral alleles. *Theor Popul Biol* **3**: 87.

Fisher RA (1932). *Statistical Methods for Research Workers*. Oliver and Boyd: Edinburgh.

Garrigan D, Hedrick PW (2003). Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* **57**: 1707–1722.

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000). SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Res* **10**: 1579–1586.

Goudet J (2000). FSTAT, a programme to estimate and test gene diversities and fixation indices (version 2.9.1). <http://www.unil.ch/izea/software/fstat.html>.

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

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

Hedrick PW (2000). *Genetics of Populations*, 2nd edn. Jones and Bartlett: Boston, MA.



- Hedrick PW, Thomson G (1983). Evidence for balancing selection at Hla. *Genetics* **104**: 449–456.
- Hedrick PW, Thomson G, Klitz W (1987). Evolutionary genetics and Hla – another classic example. *Biol J Linn Soc* **31**: 311–331.
- Hewitt G (2000). The genetic legacy of the quaternary ice ages. *Nature* **405**: 907–913.
- Hill AVS, Yates SNR, Allsopp CEM, Gupta S, Gilbert SC, Lalvani A et al. (1994). Human-leukocyte antigens and natural-selection by malaria. *Philos Trans R Soc Lond B Biol Sci* **346**: 379–385.
- Hurlbert SH (1971). Nonconcept of species diversity – critique and alternative parameters. *Ecology* **52**: 577.
- Jordan WC, Bruford MW (1998). New perspectives on mate choice and the MHC. *Heredity* **81**: 127–133.
- Kemp SJ, Hishida O, Wambugu J, Rink A, Longeri ML, Ma RZ et al. (1995). A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Anim Genet* **26**: 299–306.
- Klein J (1987). Origin of major histocompatibility complex polymorphism – the transspecies hypothesis. *Hum Immunol* **19**: 155–162.
- Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G (2003). PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. In: Altman RB et al. (eds). *Pacific Symposium on Biocomputing 8*. World Scientific: Singapore. pp 514–525.
- Markow T, Hedrick PW, Zuerlein K, Danilovs J, Martin J, Vyvilt T et al. (1993). Hla polymorphism in the havasupai – evidence for balancing selection. *Am J Hum Genet* **53**: 943–952.
- Moghaddam PH, de Knijff P, Schipper RF, Moghaddam PH, Kazemi M, Lambooj SLE et al. (1998). Selective co-evolution of the D6STNFA microsatellite region with HLA class I and II loci. *Tissue Antigens* **52**: 213–219.
- Paterson S (1998). Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *J Hered* **89**: 289–294.
- Raymond M, Rousset F (1995). Genepop (version-1.2) – population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Ryder ML (1984). Sheep. In: Mason IL (ed). *Evolution of Domesticated Animals*. Longman: London and New York. pp 63–85.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory press: New York.
- Schlotterer C, Kauer M, Dieringer D (2004). Allele excess at neutrally evolving microsatellites and the implications for tests of neutrality. *Proc R Soc Lond B Biol Sci* **271**: 869–874.
- Schwaiger FW, Epplen JT (1995). Exonic Mhc-Drb polymorphisms and intronic simple repeat sequences – janus faces of DNA-sequence evolution. *Immunol Rev* **143**: 199–224.
- Schwaiger FW, Buitkamp J, Weyers E, Epplen JT (1993). Typing of artiodactyl Mhc-Drb genes with the help of intronic simple repeated Dna-sequences. *Mol Ecol* **2**: 55–59.
- Schwaiger FW, Weyers E, Buitkamp J, Ede AJ, Crawford A, Epplen JT (1994). Interdependent Mhc-Drb exon-plus-intron evolution in artiodactyls. *Mol Biol Evol* **11**: 239–249.
- Slade RW, McCallum HI (1992). Overdominant vs frequency-dependent selection at MHC loci. *Genetics* **132**: 861–862.
- Slatkin M (1994). An exact test for neutrality based on the ewens sampling distribution. *Genetical Research* **64**: 71–74.
- Steffen P, Eggen A, Dietz AB, Womack JE, Stranzinger G, Fries R (1993). Isolation and mapping of polymorphic microsatellites in cattle. *Anim Genet* **24**: 121–124.
- Takahata N, Nei M (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967–978.
- van Haeringen WA, Gwakisa PS, Mikko S, Eythorsdottir E, Helm LE, Olsaker I et al. (1999). Heterozygosity excess at the cattle DRB locus revealed by large scale genotyping of two closely linked microsatellites. *Anim Genet* **30**: 169–176.
- Watterson GA (1978). Homozygosity test of neutrality. *Genetics* **88**: 405–417.
- Worley K, Carey J, Veitch A, Coltman DW (2006). Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*). *Mol Ecol* **15**: 623–637.
- Zhang D-X, Hewitt GM (1998). Isolation of DNA from preserved specimens. In: A Karp, PG Isaac, DS Ingram (eds). *Molecular Tools for Screening Biodiversity: Plants and Animals*. Chapman and Hall: London. pp 10–23.