Molecular Ecology (2008) **17**, 4053–4067 doi: 10.1111/j.1365-294X.2008.03892.x

Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois

S. MONA,* B. CRESTANELLO,*† S. BANKHEAD-DRONNET,* E. PECCHIOLI,† S. INGROSSO,‡ S. D'AMELIO,‡ L. ROSSI,§ P. G. MENEGUZ§ and G. BERTORELLE*

*Department of Biology and Evolution, University of Ferrara, Via Borsari 46, 44100 Ferrara, Italy, †Centre of Alpine Ecology, Viote del Monte Bondone, 38040 Trento, Italy, ‡Department of Public Health Sciences, University of Rome 'La Sapienza', P.le Aldo Moro 5, 00185 Rome, Italy, §Department of Animal Production, Epidemiology and Ecology, University of Turin, Via Leonardo da Vinci 44, 10095 Grugliasco (TO), Italy

Abstract

The major histocompatibility complex (MHC) harbours some of the most polymorphic loci in vertebrate genomes. MHC genes are thought to be subject to some form of balancing selection, most likely pathogen-mediated selection. Hence, MHC genes are excellent candidates for exploring adaptive processes. In this study, we investigated the genetic variation at exon 2 of the DRB class II MHC locus in 191 alpine chamois (*Rupicapra rupicapra*) from 10 populations in the eastern Alps of Italy. In particular, we were interested in distinguishing and estimating the relative impact of selective and demographic factors, while taking into account the confounding effect of recombination. The extremely high d_n/d_s ratio and the presence of trans-species polymorphisms suggest that a strong long-term balancing selection effect has been operating at this locus throughout the evolutionary history of this species. We analysed patterns of genetic variation within and between populations, and the mitochondrial D-loop polymorphism patterns were analysed to provide a baseline indicator of the effects of demographic processes. These analyses showed that (i) the chamois experienced a demographic decline in the last 5000–30 000 years, most likely related to the postglacial elevation in temperature; (ii) this demographic process can explain the results of neutrality tests applied to MHC variation within populations, but cannot justify the much weaker divergence between populations implied by MHC as opposed to mitochondrial DNA; (iii) similar sets of divergent alleles are probably maintained with similar frequencies by balancing selection in different populations, and this mechanism is also operating in small isolated populations, which are strongly affected by drift.

Keywords: balancing selection, bottleneck, MHC, mtDNA, population structure, recombination, *Rupicapra rupicapra*

Received 9 April 2008; revision received 4 July 2008; accepted 6 July 2008

Introduction

The analysis of patterns and levels of genetic variation at neutral markers, such as autosomal microsatellites and mitochondrial DNA regions, has been widely used in the last decades to infer historical events (e.g. past demographic expansions or contractions) and geographical features (e.g. fragmentation) in natural populations (Hartl & Clark 2006). However, the study of molecular polymorphisms at loci under selection is the only direct way to understand the genetics of adaptive processes (Hedrick 2001). This is a fundamental issue in evolutionary studies. Recent and rapid environmental changes due to human activities further underline the importance of studying genetic adaptation in conservation biology (Gutierrez-Espeleta *et al*. 2000; van der Walt *et al*. 2001; Aguilar *et al*. 2004).

Correspondence: Giorgio Bertorelle, Fax: +390532249771; E-mail: ggb@unife.it

A classical example of a genetic region where adaptive variation has been analysed is the major histocompatibility complex (MHC) (Potts & Wakeland 1990; Hedrick 1994). The MHC is a multigene family encoding receptors involved in the immune response of vertebrates (Trowsdale 1993). MHC molecules bind fragments of foreign proteins displaying them on the cell surface. This complex can be recognized by T-cells, initiating a cascade of immune response (Ploegh & Watts 1998). MHC includes two major subfamilies, including the class I and class II genes. Class I genes are expressed on the surface of almost all nucleated somatic cells (Piertney & Oliver 2006) and recognize endogenously derived peptides. Hence, they are primarily devoted to responding to intracellular pathogens such as viruses (Klein 1986). Class II genes are expressed only on antigen-presenting cells and are responsible of monitoring the extracellular environment, thus providing protection from parasites and bacteria (Klein 1986).

The MHC is probably the most polymorphic region in the nuclear genome of vertebrates (Robinson *et al*. 2003), displaying an unusually large number of alleles and high nucleotide diversity compared to other loci (Parham & Ohta 1996). The abundant polymorphism observed at the MHC is generally attributed to some form of balancing selection. This is confirmed by empirical evidence on variation mostly concentrated in codons involved in the recognition of foreign peptides (the peptide-binding region, hereafter named PBR) for both class I and II genes (Hughes & Yeager 1998). It is usually assumed that selection operating on the MHC is pathogen-mediated (Doherty & Zinkernagel 1975; Apanius *et al*. 1997; Hedrick & Kim 2000). Another mechanism proposed to explain the high levels of polymorphism at the MHC is sexual selection (Jordan & Bruford 1998; Penn & Potts 1999). A number of studies have shown that MHC can influence mate choice in natural populations (Bonneaud *et al*. 2006; Schwensow *et al*. 2007a; see Bernatchez & Landry 2003 for a review). Pathogen-mediated and sexual selection mechanisms are not necessarily mutually exclusive. They can be reconciled under the hypothesis that mating preferences for MHC genotypes evolved to guarantee parasite resistance to offspring (Piertney & Oliver 2006).

Overdominance and frequency-dependent selection are two of the main pathogen-mediated balancing selection processes that have been suggested to explain genetic diversity patterns observed at most MHC loci (see reviews of Meyer & Thomson 2001; Bernatchez & Landry 2003; Garrigan & Hedrick 2003; Sommer 2005; Piertney & Oliver 2006). The overdominance hypothesis implies that heterozygous individuals have increased fitness with respect to homozygotes. This advantage can be symmetric, where all heterozygotes have equal fitness (Doherty & Zinkernagel 1975), or asymmetric, with a more pronounced advantage for heterozygotes of divergent alleles (Wakeland *et al*. 1990). On the other hand, the frequency-dependent selection hypothesis proposes an advantage for rare alleles, which may occur when the majority of pathogens are adapted to escaping common MHC proteins (Takahata & Nei 1990). This selective mechanism would eventually produce a dynamic system of host–parasite co-evolution, where the fitness of each allele varies cyclically through time.

In principle, several tests and analyses can be used to detect selection operating at the MHC, and, if this is the case, to determine the extent, the temporal scale and the most likely selective mechanism involved (Garrigan & Hedrick 2003; Sommer 2005; Piertney & Oliver 2006). Genetic variation within and between species is commonly used for this purpose. However, at least three factors should be considered, which may interact with selection to confound the results: (i) intragenic recombination, which variously modifies measures of genetic variation used to test the null hypothesis of neutral evolution and affects the interpretation of gene genealogy (Wall 1999; Schierup *et al*. 2001; Richman *et al*. 2003a); (ii) demographic history, which leads to population size variations that can produce deviations from the neutral expectation similar to that observed under selection (Kreitman 2000; Otto 2000; Nielsen 2005); (iii) population structure, since the pathogen community at a specific time may vary among populations and, consequently, the most advantageous MHC alleles may also vary among host populations (Nevo & Beiles 1992; Muirhead 2001; Hedrick 2002). In this study, we analysed genetic variability at exon 2 of the MHC class II DRB locus in the Alpine chamois (*Rupicapra rupicapra rupicapra*), paying particular attention to these three factors. The pattern of MHC variation was analysed using classical methods. We also (i) estimated the impact of recombination on MHC variation and modified accordingly the significance levels of some of the statistical tests used to detect selection; (ii) considered the genetic variation observed at the mitochondrial DNA (mtDNA) control region, used here as a neutral control to distinguish between demographic and selective processes; and, finally, (iii) analysed the population structure at this gene on the basis of 10 populations located in the eastern Italian Alps.

Materials and methods

Sample collection

Chamois (genus *Rupicapra*, subfamily Caprinae) are mountain ungulates inhabiting medium- to high-altitude mountain ranges of several regions in southern Europe, including the Carpathians and the Middle East. Tissue samples (skeletal muscle and skin) were collected between 1997 and 2005 from 191 Alpine chamois (*Rupicapra rupicapra rupicapra*) from 10 populations inhabiting the the eastern Alps in the provinces of Trento, Bolzano and Belluno of Italy (see Fig. 1 and Table 1).

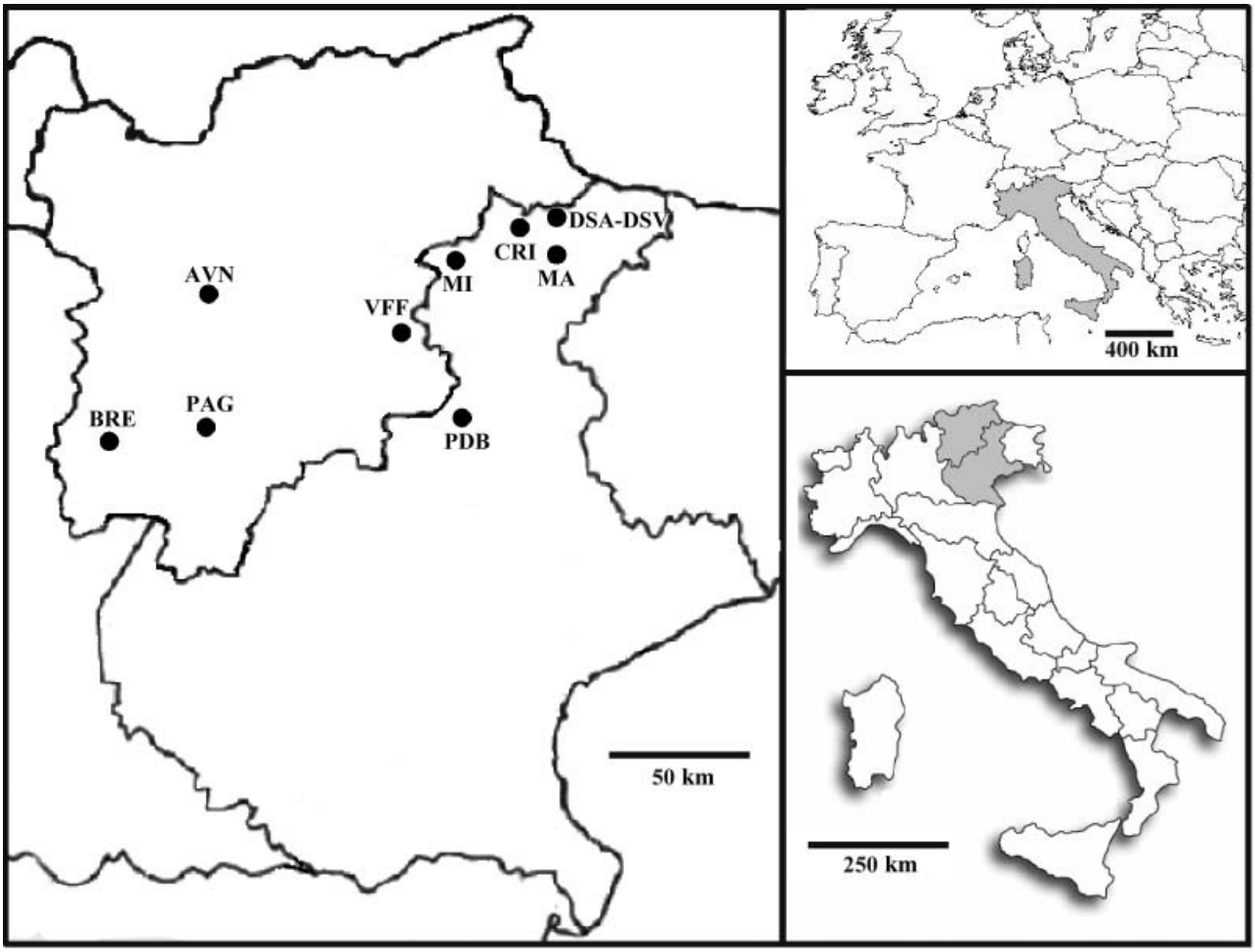

Fig. 1 Map of the Eastern Alps of Italy showing the location of the 10 alpine chamois populations analysed. Population codes are as in Table 1. Black lines indicate the borders between countries or Italian regions.

Laboratory analyses

Genomic DNA was extracted from frozen (–80 °C) or alcohol-preserved (95% ethanol) tissues using the DNEasy Tissue Kit (QIAGEN). Polymerase chain reaction (PCR) amplification of the MHC class II DRB gene (231-bp fragment of exon 2) was performed using the two cattle primers HL030, HL031 and a nested primer HL032 as described in Schaschl *et al*. (2004). Each amplicon was purified using ExoSAP-IT (Amersham Biosciences) and sequenced in both directions with BigDye Terminator Kit version 1.1 (Applied Biosystems). Sequencing products were cleaned by centrifugation through Sephadex G-50 (Sigma-Aldrich) and run on an ABI PRISM 3100 sequencer (Applied Biosystems).

The complete mtDNA control region was amplified by PCR using the primer MF (Mannen *et al*. 2001) and Hphe (Douzery & Randi 1997). Two microlitres (approximately 100 ng) of DNA was used as template for amplification in a 20-μL reaction mix containing: *Taq* buffer 1× (Polymed), 3 mM $MgCl_2$, 100 μM dNTPs, 25 μM of each primer, and 1 U of *Taq*. The thermocycling regime consisted of incubation at 94 °C for 2 min, followed by 35 cycles of 94 °C for 15 s, 60 °C for 1 min and 72 °C for 1 min, with a final extension of

Table 1 Sample sizes for DNA regions. Asterisks indicate the samples typed for the entire mtDNA control region; the other samples were typed only at the first 360 bp

Population	Population Code	MHC	mtDNA
Val Marzon (Dolomiti di Sesto)	DSV	26	20
Auronzo (Dolomiti di Sesto)	DSA	26	11
Monte Cristallo	CRI	9	4
Marmarole	MA	30	14
Migon	MI	16	15
Parco Dolomiti Bellunesi	PDB	10	2
Breguzzo	BRE	18	28*
Paganella	PAG	20	30*
Val di Fiemme e Fassa	VFF	19	49*
Alta Val di Non	AVN	17	19*

72 °C for 5 min. Each amplicon was purified using ExoSAP-IT (Amersham Biosciences) and the sequence of approximately the first 360 bp of the control region was determined using the primer MF and H493 (Douzery & Randi 1997). Four out of 10 chamois populations considered here were also typed for the entire control region, using the additional primer L362 (Douzery & Randi 1997). The sequencing programme consisted of 25 cycles at 96 °C for 10 s, 60 °C for 5 s, and 60 °C for 1.5 min. Sequencing products were cleaned by centrifugation through Sephadex G-50 (Sigma-Aldrich) and run on an ABI PRISM 3100 sequencer (Applied Biosystems). The resulting sequences were edited with Chromas version 1.45 (www.technelysium.com.au/chromas.html), aligned using Clustal_X (Thompson *et al*. 1997) and checked by eye. In Table 1, we reported the sample sizes for each population and each marker, including the length of the DNA fragments. All DRB exon 2 alleles and the new mtDNA sequences were deposited in GenBank under the accession nos EU887297–EU887509.

Statistical analyses

MHC haplotype reconstruction. The haplotypes of each individual were inferred using the Bayesian approach implemented in Phase 2.1 (Stephens *et al*. 2001; Stephens & Donnelly 2003). Intragenic recombination has been found in MHC class II genes of various species (Gyllensten *et al*. 1991; Richman *et al*. 2003b; Schwensow *et al*. 2007b), including the chamois (Schaschl *et al*. 2005). For this reason, haplotypes were reconstructed under the approximate coalescent model with variable recombination rate described in Li & Stephens (2003). The posterior probabilities of each haplotype were approximated through 1 000 000 Markov chain Monte Carlo (MCMC) iteration with a 10% burn-in period. The analysis was run twice and the results were compared both by visually examining the traces of the continuous parameters and the marginal likelihood and by looking at the posterior probability of the inferred haplotypes for each individual. The reconstruction of the alleles in nuclear multiple heterozygous individuals using statistical algorithms is now a common procedure for many genes, and it has also been applied in several MHC/human leukocyte antigen (HLA) studies (Hosomichi *et al*. 2006; Meyer *et al*. 2006; Hess *et al*. 2007; Sanchez-Mazas 2007). Among these methods, the high performance of Phase has been demonstrated both using real (including MHC sequences) and simulated data (Stephens & Donnelly 2003; Marchini *et al*. 2006; Bos *et al*. 2007).

Recombination rate estimation. The population recombination parameter ($\rho = 4N_e r$) was estimated using the composite-likelihood approach of Hudson (2001) extended by McVean *et al*. (2002) to account for recurrent mutations, as implemented in the software LDhat (McVean *et al*. 2002). We first computed an approximate finite sites version of the population mutation rate parameter ($\theta = 4N_e \mu$) and then used LDhat to estimate the likelihood of the observed configuration of each pair of segregating sites given θ for values of ρ ranging from $0 \leq \rho \leq 50$. The recombination rate was then obtained by maximizing the likelihoods from all pairwise comparisons. The presence of recombination was also tested using the permutation approach implemented in LDhat (McVean *et al*. 2002).

Nonsynonymous vs. synonymous substitutions. The ratio of nonsynonymous (d_n) to synonymous (d_s) rates of substitution was computed using the modified Nei & Gojobori distance (1986) applying the Jukes–Cantor correction for multiple hits with the software MEGA 3.1 (Kumar *et al*. 2004). The transition/transversion parameter in the modified Nei and Gojobori distance (Nei & Kumar 2000) was determined in analyzing the exon 2 allelic genealogy with the HKY + Γ model. The d_n and d_s rates were computed in the total sample weighting them by the frequencies of each allele pair (Hedrick *et al*. 2000); the significance of their difference was computed using the *Z*-test of selection (with the variance estimated after 1000 bootstrap replicates) under the null hypothesis of $d_n = d_s$.

Trans-species polymorphisms. The possible occurrence of trans-species polymorphisms (Klein 1980) was analysed considering the DRB exon 2 sequence of various species belonging to the Caprinae subfamily, and adding a *Bos taurus* sequence as outgroup. The complete list of the species and the alleles is available in Table S1, Supporting Information. The phylogenetic tree was reconstructed using MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003), with the model of nucletotide substitution selected by MrModelTest (Nylander 2004) according to the Akaike information criterion (AIC). The analysis was the result of two runs of four incrementally heated chains with default temperature and 20 000 000 generations long, with a 30% burn-in. Convergence was checked by examining the generation plot visualized with Tracer (Rambaut & Drummond 2004) and computing the potential scale reduction factor (Gelman & Rubin 1992). This statistic is based on the assumption that, at convergence, the within-chain variation should be indistinguishable from the between-chain variation, while before convergence the latter should be greater than the former.

Neutrality tests. Departures from neutrality were evaluated using three statistical tests, as implemented in Arlequin 3.11 (Excoffier *et al*. 2005): Tajima's *D* (Tajima 1989), Fu's *Fs* (1997), and the Ewens–Watterson–Slatkin exact test (Watterson 1978; Slatkin 1994; Slatkin 1996). The first two tests are based on the site frequency spectrum, and are therefore influenced by long-term mutational patterns, as well as on more recent population dynamics (Garrigan & Hedrick 2003). In contrast, the third test is based on allele frequencies only and does not

consider the molecular difference between alleles. Therefore, it is most sensitive to recent population dynamics which produce a departure from the probability distribution of the sample configuration expected under neutrality. The statistical significance for Tajima's *D* and Fu's *Fs* was determined by generating their null distribution under a coalescent neutral model with recombination with the software MLCOALSIM (Ramos-Onsins & Mithell-Olds 2007). We performed 2000 coalescent replicates for each population using the estimated population mutation rate parameter and the maximum likelihood estimate of the population recombination parameter (see above). The Ewens–Watterson–Slatkin exact test is insensitive to recombination (Zeng *et al*. 2007), and the *P* values were therefore computed using 2000 replicates without taking recombination into account.

Demographic inferences using the mtDNA. The demography of the chamois populations was initially investigated using the Bayesian skyline plot (Drummond *et al*. 2005) in the four populations for which the complete mtDNA control region was available. The skyline plot is a nonparametric estimate of the changes in population size that makes no a priori assumptions on the demographic model. In other words, this analysis reconstructs the variation of the effective population size through time, without the need to specify a priori, for example, if the population increased (or decreased) exponentially or followed a sudden expansion (or contraction) model (Pybus *et al*. 2000). In practice, it is necessary to define a number of groups of internodes sharing the same effective population size, but the sizes of each of these groups as well as the gene genealogy (and all other relevant mutational parameters) are sampled in a Monte Carlo Markov chain framework (Drummond *et al*. 2005). We ran each analysis twice using the software BEAST (Drummond & Rambaut 2007), for 10 000 000 generations each time, with a 10% burn-in, dividing the internodes into three groups. The skyline reconstruction was performed with Tracer (Rambaut & Drummond 2004), in order to visualize the effective population size as a function of time. To express time in years, we first determined the mutation rate for the mitochondrial control region before running the skyline analysis. Mutation rate estimates can be obtained calibrating the molecular clock over known divergence times between two or more taxa, but all sources of uncertainties related to the calibration point and the phylogenetic reconstruction should be considered in this analysis (Nei & Kumar 2000). The software BEAST was initially used to estimate the posterior distribution of the mutation rate, defining a normal prior distribution for the divergence age between *R. rupicapra* and *Rupicapra pyrenaica* with mean and standard deviation equal to 1.6 and 0.3 million years, respectively. These values were recently estimated by Lalueza-Fox *et al*. (2005) using cytochrome *b* sequences. The Yule pure birth process was used as a prior for the topology and branch lengths (Rannala & Yang 1996). The analysis was run twice, for 20 000 000 generations each, with a 10% burn-in, assuming the nucleotide substitution selected by MrModeltest under the AIC criterion. The posterior distribution of the mutation rate obtained in that way was subsequently used as a prior in the Bayesian skyline reconstruction.

BEAST was also used to estimate the posterior distribution of the growth rate, assuming a model of continuous and exponential variation of the population size through the entire history of the sampled sequences.

Population divergence and simulations. The level of divergence between populations at both markers (MHC and mtDNA) was estimated using the analysis of molecular variance (Excoffier *et al*. 1992), as implemented in Arlequin 3.11 (Excoffier *et al*. 2005). F_{ST} values (also called Φ_{ST}) were computed using the number of differences as a molecular distance between alleles (different metrics did not modify the results). Standard errors were based on 1000 bootstrap replicates.

A direct comparison between the F_{ST} values at mitochondrial and nuclear markers is not possible, since the effective population size at the former is four times smaller than the effective size at the latter. In order to understand whether any difference between the population divergence at the two markers can be attributed to the different effective sizes only, or if some additional process (i.e. selection) should be assumed at the MHC, we performed a small simulation experiment. We assumed a nonequilibrium model with eight populations. In the simulation, the populations diverged from a common ancestral group 3500 generations ago (compatible with the splitting of our chamois population approximately 20 000 years ago and a generation time of 6.2 years (Pérez *et al*. 2002)), decreased in size and exchanged only few migrants since then. The model was simulated using the software SimCoal2 (Laval & Excoffier 2004), with the following parameters: effective size of the ancestral group: 300; migration rate after the split: 0.0004; present-day effective size for each populations: 100; length of the simulated sequence: 360 bp; mutation rate: 0.00002 (per generation per locus). The parameter values were defined with the goal of obtaining a final group of populations having a level of genetic variation and divergence similar to what was observed at the mtDNA in the chamois. The simulations were not performed to systematically explore all the possible scenarios. We repeated the simulations to estimate the distribution of the F_{ST} values expected under this model for a nuclear marker with sequence characteristics similar to our MHC sequences. We used the same demographic parameters, with population size four times that of mtDNA, a mutation rate of 1.3×10^{-6} per locus per generation (compatible with that estimated for MHC loci; Satta *et al*. 1991), and a sequence of 231 bp. Finally, we reconstructed the empirical frequency

Table 2 Asterisks indicate the samples showing a statistically significant value of ρ ($P < 0.001$), as determined by the permutation approach implemented in LDhat

	n	θ	ρ	ρ/θ
DSV	26	4.87	8.84*	1.81
DSA	26	4.63	12.50*	2.70
CRI	9	5.76	5.68*	0.99
MA	30	4.72	6.32*	1.34
MI	16	5.47	6.32*	1.16
PDB	10	4.37	0.00	0.00
BRE	18	4.26	4.42*	1.04
PAG	20	4.41	6.95*	1.58
VFF	19	4.98	5.68*	1.14
AVN	17	5.65	9.47*	1.68

n, sample size; θ, finite sites population mutation parameter; ρ, maximum likelihood estimate of population recombination parameter.

distribution of the ratio between the F_{ST} values from the two sets of simulations.

Results

MHC DRB exon 2 variability in the alpine chamois and intragenic recombination

The sequence of 231 bp (77 codons) at the *DRB* exon 2 locus was obtained for all the 191 individuals. All the individuals with heterozygous sites producing different alleles were typed independently in two different laboratories (from the DNA extraction to the sequence), obtaining consistent results. No evidence of multiple locus amplification was found in the electropherograms, confirming previous reports within the genus *Rupicapra* (Schaschl *et al*. 2005; Alvarez-Busto *et al*. 2007). A total of 28 alleles (defined by 22 segregating sites) were identified (see Fig. S1), seven of which are identical to alleles previously described by Schaschl *et al*. (2005) in the Austrian population and representing 85.6% of our sample. The mean pairwise difference between alleles was 8.50 and 6.80 for nucleotide and amino acid sites, respectively.

The new haplotypes were coded from Ruru-DRB20 to Ruru-DRB40 in agreement with the nomenclature proposed by Klein *et al*. (1990), starting from the last haplotype identified in Schaschl *et al*. (2005). The allele frequencies observed in each population are available in Table S2, Supporting Information. We found 53 homozygotes among 191 individuals. Significant deviation from Hardy–Weinberg expectation was observed only in one population (AVN, where the observed heterozygosity was lower than expected). The large number of homozygotes and the absence of Hardy–Weinberg deviations (with a single significant deficit of heterozygotes) provide additional evidence that only one locus was amplified. Following the model proposed by Brown *et al*. (1993) for the DRB exon 2 protein structure in humans, we attributed 21 of the 77 codons to the PBR. Almost 82% of the polymorphic sites (18/22) fell in the PBR.

In all the samples separately analysed, with the exception of the small group of individuals from the Dolomiti Bellunesi National Park (PDB, $n = 10$), recombination was significantly different from zero ($P < 0.001$) (Table 2). The ρ/θ ratio was always close to 1 (Table 2), showing that the two forces (recombination and mutation) had similar weights in driving the evolution of this locus in the chamois.

Nonsynonymous vs. synonymous substitutions

The sequences we analysed showed no synonymous substitutions. The ratio d_n/d_s was not infinity since d_s could be estimated from the codons displaying more than one variable site, but it was nonetheless very large at 54.8 for the entire sequence and 42.4 for the PBR region. These values are larger than any other ratio previously reported for ungulates (see Bernatchez & Landry 2003, and references therein; Table 2 in Amills *et al*. 2004; Schaschl *et al*. 2004; Alvarez-Busto *et al*. 2007). The rate of nonsynonymous substitutions, d_n, was much higher in the PBR region compared to the whole sequence ($d_n = 0.122 \pm 0.035$ vs. $d_n = 0.040 \pm 0.011$). Conversely, d_n/d_s was lower in the PBR region than in the complete sequence. This result seems counterintuitive, but it can be easily explained by the fact that all the codons with more than one variable site were in the PBR region, hence determining a higher d_s than that estimated in the whole sequence. In principle, the absence of synonymous substitutions invalidates the assumption of the *Z*-test used to test the difference between d_n and d_s rates (Nei & Kumar 2000), and the *P* values we found ($P = 0.00040$ and $P = 0.00044$ for the entire region and PBR only, respectively) might be therefore biased. At the same time, however, the absence of synonymous substitutions argues convincingly for a strong positive selection process acting on this locus.

Trans-species polymorphisms

The Bayesian analysis of interspecific phylogeny was performed assuming the GTR + Γ + I model of nucleotide substitution. This was the most appropriate model selected by MrModeltest. The resulting tree is presented using the 50% majority consensus rule (Fig. 2). The tree is not fully resolved, but trans-species polymorphisms can be clearly detected in the Caprinae subfamily. Considering that the Caprinae radiation occurred in the late Miocece (Lalueza-Fox *et al*. 2005), this result implies that similar alleles have been maintained for at least 5 million years ago.

Lineages corresponding to the two *Rupicapra* species (*R. rupicapra* and *R. pyrenaica*) are not only intermixed in the MHC phylogeny and in its highly supported subclades,

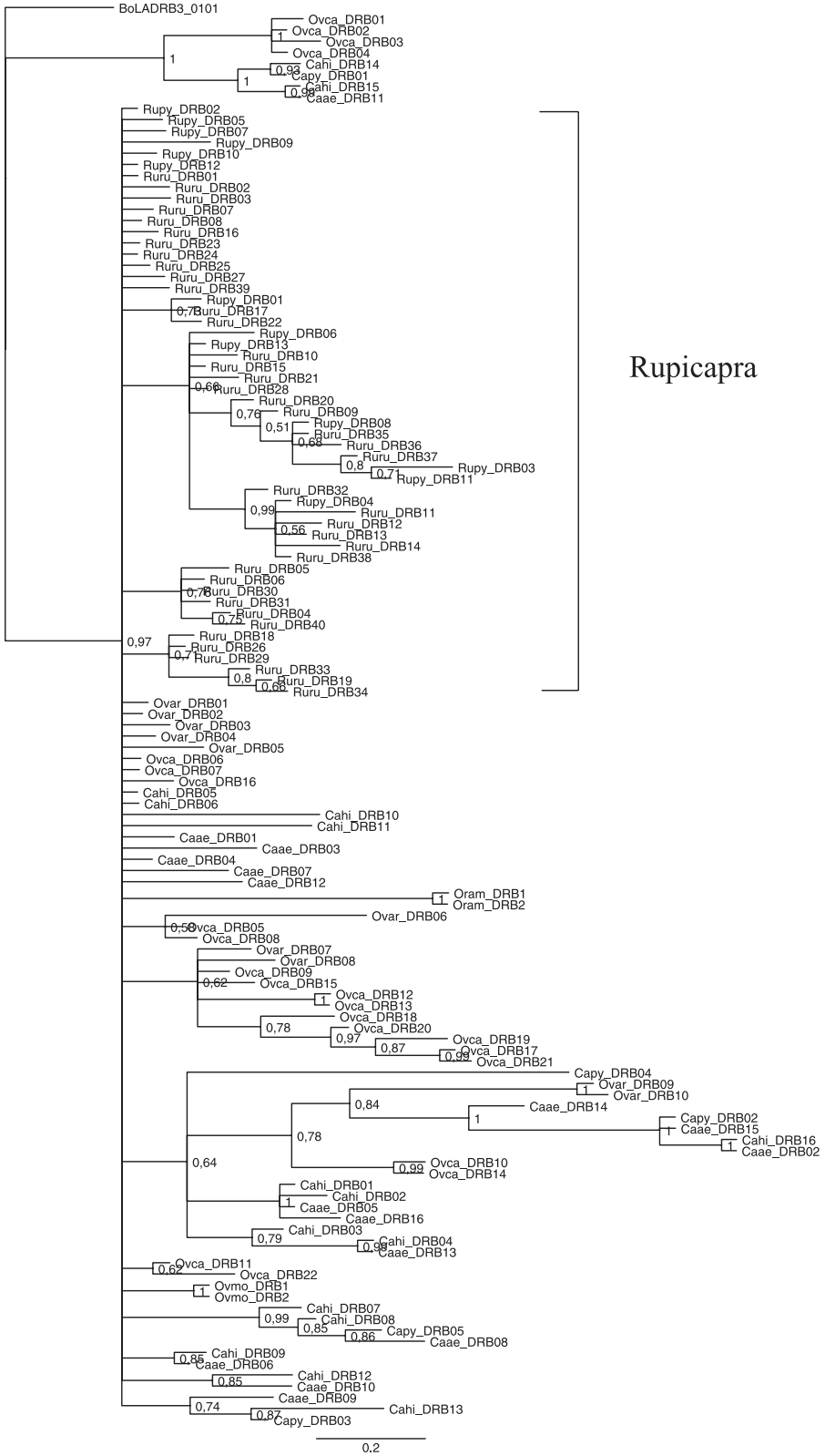

Fig. 2 Fifty per cent majority consensus tree of MHC haplotypes generated using MrBayes under a GTR + Γ + I nucleotide substitution model. Tree was rooted using a *Bos taurus* sequence as outgroup (BolaDRB3_0101). Ruru and Rupy correspond to haplotypes observed in the Alpine (*Rupicapra rupicapra*) and the Pyrenean chamois (*Rupicapra pyrenaica*), respectively. A complete list of the sequences analysed with their ID numbers is available in Table S1. Posterior probabilities higher than 0.50 are shown.

	D	*D P* val (no rec)	*D P* val (rec)	*Fs*	*Fs P* val (no rec)	*Fs P* val (rec)	E–W–S *P* val	H–W *P* val
DSV	–0.107	NS	NS	2.183	NS	**	NS	NS
DSA	0.018	NS	NS	2.138	NS	**	NS	NS
CRI	0.789	NS	NS	3.515	*	**	NS	NS
MA	1.756	*	**	5.688	NS	**	NS	NS
MI	1.346	NS	NS	3.079	NS	*	NS	NS
PDB	0.035	NS	NS	4.761	*	*	NS	NS
BRE	2.766	**	**	2.871	NS	*	NS	NS
PAG	2.575	**	**	3.348	NS	**	NS	NS
VFF	1.292	NS	*	4.295	NS	**	NS	NS
AVN	1.490	*	*	0.561	NS	NS	*	*

*indicates *P* values -0.05; **indicates *P* values ≤ 0.01; NS, *P* > 0.05.

Table 3 Tajima's *D*, Fu's *Fs*, Ewens–Watterson–Slatkin (E–W–S) exact tests and Hardy–Weinberg equilibrium test computed for the 10 chamois populations at the MHC locus. Statistical significance was determined both under a neutral coalescent model with no recombination (no rec) and under a neutral coalescent model with recombination (rec) as explained in the text

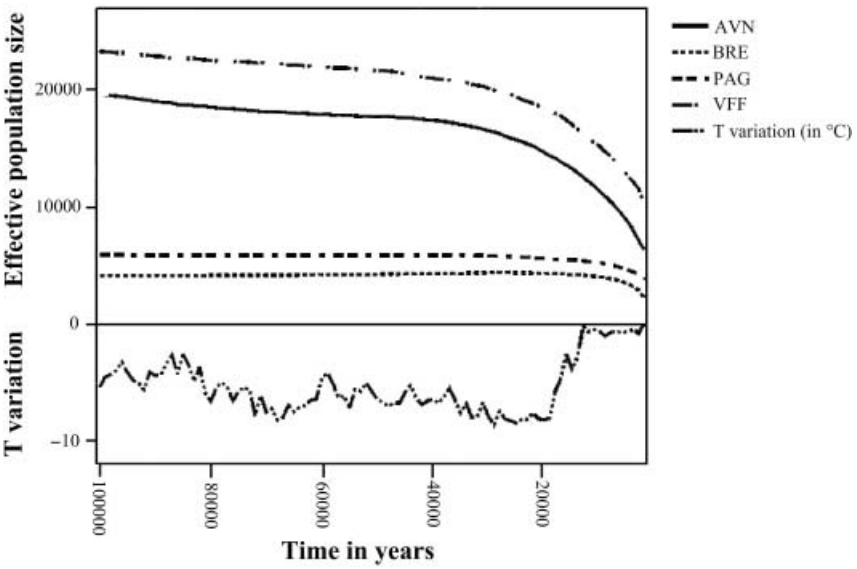

Fig. 3 Bayesian skyline plots in the four populations where the complete mtDNA control region was available (only the median of the effective size for the four populations is shown). The temperature variation (×10³) estimated from the Vostok ice core in Antarctica are also reported on the bottom line. The time is in years ago.

but some alleles are also shared among species. In other words, these alleles persisted over a time period of about 1.6 million years, which corresponds to the estimated divergence time between the Alpine and the Pyrenean chamois (Lalueza-Fox *et al*. 2005).

Neutrality tests

All but one population (AVN for Tajima's *D*) showed positive values for both Tajima's *D* and Fu's *Fs* neutrality tests (Table 3). Positive values are consistent with internal population structure, balancing selection or a demographic bottleneck (Tajima 1989; Charlesworth *et al*. 1993; Kreitman 2000), but intragenic recombination modifies the frequency distribution of these statistics under neutrality (Wall 1999; Ramos-Onsins & Rozas 2002). The statistical significance of the observed *D*'s and *Fs*'s was thus computed generating the null distribution of both *D* and *Fs* under a neutral coalescent model with recombination, assuming the recombination rate previously estimated. When recombination was considered, the total number of significant tests increased from 6 to 14 out of 20 (Table 3); Tajma's *D* was significant with $0.01 < P < 0.05$ and $P < 0.01$ in two and three populations, respectively, and Fu's *F* was significant with $0.01 < P < 0.05$ and $P < 0.01$ in three and six populations, respectively (Table 3). On the contrary, only one out of 10 populations displayed a significant departure from the neutral expectation when the Ewens–Watterson–Slatkin test was applied.

Demographic inferences using the mtDNA

When the demographic history of the chamois was analysed in four populations using the whole mtDNA control region, a simple model of constant population size appeared unlikely. The Bayesian skyline plots for AVN and VFF support a demographic decline which accelerated in the last 20 000 to 30 000 years (Fig. 3). A different pattern was obtained for BRE and PAG, with a lower and constant effective size, and

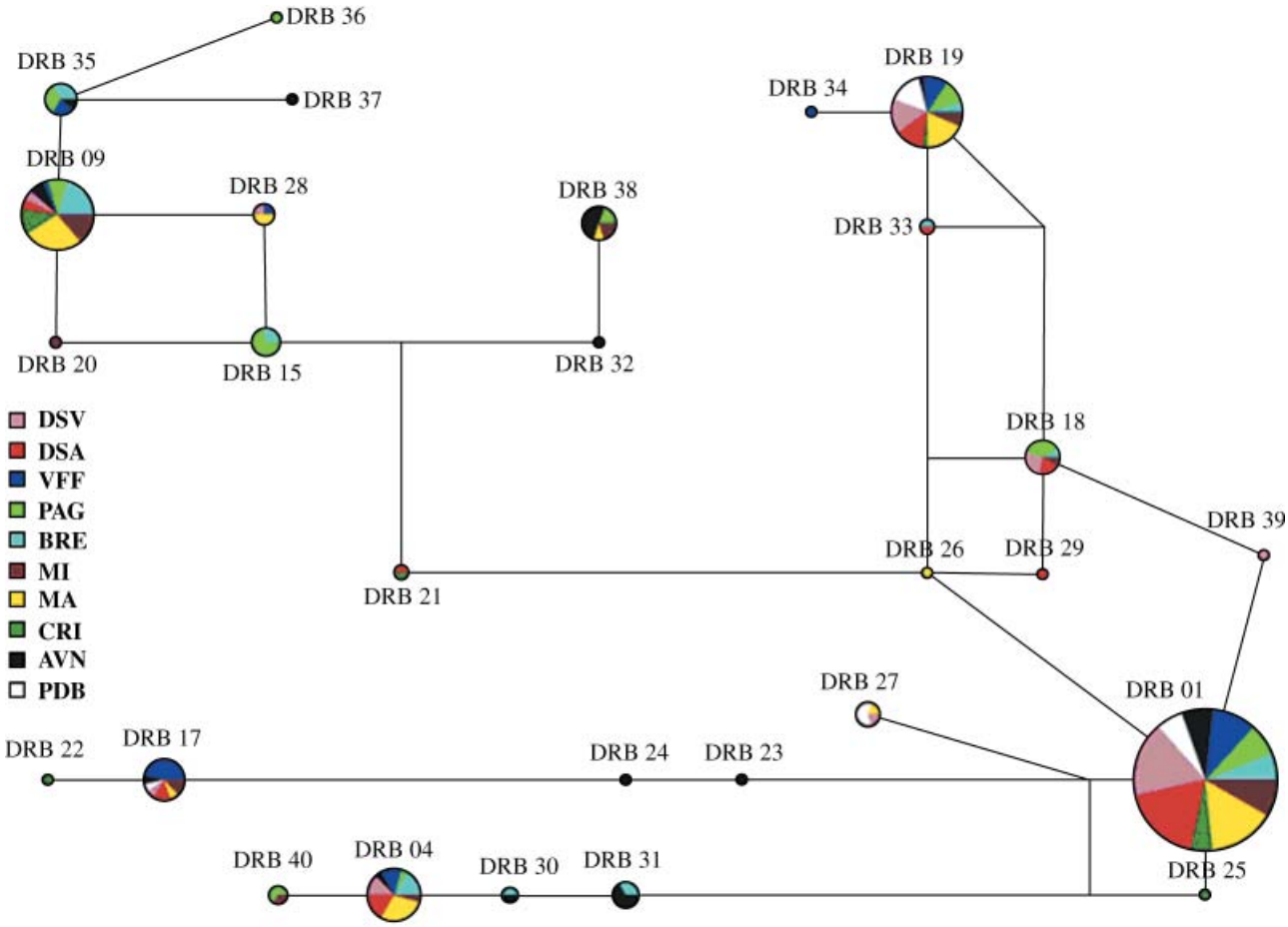

Fig. 4 Median-joining network (Bandelt *et al*. 1999) of the 28 MHC alleles found in the 10 populations considered. Branch lengths are proportional to number of mutations; node diameter is proportional to allele frequency (see also Table S2); median vectors not shown. Colours: black (AVN), light blue (BRE), green (CRI), red (DSA), pink (DSV), yellow (MA), violet (MI), light green (PAG), white (PDB), blue (VFF).

Table 4 Mean and 95% HPD of the posterior distribution of the growth rate per year (scaled by the mutation rate per site per year) estimated in the four populations for which the complete mitochondrial control region was available. Coalescent model: exponential growth; mutation model: Hasegawa–Kishino–Yano. Posterior distribution was approximated using 10 000 000 MCMC with a 10% burn-in

	n	Mean	95% lower HPD	95% upper HPD
BRE	28	–452.1	–998.3	–12.1
PAG	30	–393.2	–892.8	45.3
VFF	49	–67.3	–402.4	289.4
AVN	19	–174.1	–570.7	192.5

n denotes the sample size of the populations considered.

possibly a more recent decline in the last few thousand years ago (Fig. 3). Several factors may be responsible for the difference observed among these geographically related groups, including the confounding effects produced by recent human activities such as hunting and management. It is worth noting, however, that the inferred demographic decline, considering the likely errors associated with the mutation rate estimation, is compatible with the postglacial elevation of temperatures supported, for example, by the Vostok ice core in Antarctica (Petit *et al*. 1999; see Fig. 3). The results obtained using the Bayesian skyline plots were confirmed when the exponential growth rate was estimated assuming a demographic model of continuous change in population. In fact, the posterior probability distribution of this rate was shifted towards negative values for all the samples, and in one population the 95% support interval did not include zero (Table 4).

Population divergence and simulations

The average population divergence at MHC was low, with $F_{ST(MHC)} = 0.052$ ($P < 0.001$) and the more common alleles shared among populations (Fig. 4). When we repeated the

same analysis using the mtDNA sequences (excluding CRI and PDB for the small sample sizes; Table 1), very different results were obtained. The population divergence at this marker was about 10.7 times larger than that observed at the MHC region [$F_{ST(mtDNA)} = 0.566$, $P < 0.001$].

Very similar differentiation levels (results not reported) were obtained using the classical Wright's F_{ST}, which does not consider the molecular divergence among alleles. This index is closely related to the relative increase in gene diversity which is observed when all populations are analysed together, with respect to the average value observed within populations. For the mtDNA sequences, this increase in gene diversity was substantial (0.69 within populations on the average, and 0.92 overall), but virtually zero for the MHC locus (0.74 vs. 0.75). The same pattern of population/locus genetic structure was found analysing allele sharing, a different measure of divergence. Population-specific alleles represented only 3% of the total number of alleles at MHC, but 23% for mtDNA.

In the simulation experiment specific for the mtDNA fragment, the average F_{ST}, gene diversity and number of alleles in 1000 simulated genealogies were equal to 0.59, 0.56 and 4.63, respectively, very similar to what was observed in the real data. The same demographic scenario was then simulated for the MHC locus, and the empirical frequency distribution of the ratio between the F_{ST} values from the two set of simulations was finally reconstructed. The median (2.24) and the mean (2.49) of these distributions were greater than 1, as expected given the different effective population sizes at the two markers, but much smaller than the 10.7 times ratio observed in the real data. The 95% of the F_{ST} ratio distribution lies between 1.12 and 5.31, which excludes the observed value of this ratio with $P < 0.05$. Additional simulations with different parameter values, such as 10 times larger populations sizes or migration rates, produced concordant results. In fact, the population divergence at MHC can be driven in the simulations to values similar to those reported in the real populations, but the difference between the markers resulted always much smaller than in the real data.

Discussion

Understanding the relative role of selection, demography and migration in shaping patterns and levels of genetic variation in natural populations is not an easy task. In fact, the null hypothesis tested in many statistical techniques is violated not only by natural selection but also by demographic instability and population structure (Andolfatto 2001; Przeworski 2002; Nielsen 2005). In addition, intragenic recombination further complicates inference, because the neutral expectations used as null hypotheses are usually developed without taking recombination into account (Ewens 1972; Tajima 1989; Schierup & Hein 2000). In this study, we used both classical methods and specific analyses to understand the processes which affected genetic variation at one gene of the MHC system in the Alpine chamois, focusing on the relative roles played by selection, recombination, demography, and population structure.

The role of recombination

Previous work on the DRB exon 2 in the chamois claimed that recombination outweighs mutation by an order of magnitude in generating new allelic diversity at this locus (Schaschl *et al*. 2005; Schaschl *et al*. 2006). Our results suggest that recombination plays an important role at this gene, and should be considered in interpreting results, but the relative weights of mutation and recombination should be revised. In virtually all populations we analysed, the ratio ρ/θ was close to 1, indicating that mutation and recombination have similar effects. The discrepancy between our results and those of Schaschl *et al*. (2006) cannot be attributed to the estimator, because both studies applied the method implemented in LDhat (McVean *et al*. 2002). However, Schaschl *et al*. (2006), computed ρ on the basis of different sequences (haplotypes) only and not using all observed sequences as required by a coalescent-based method. In other words, we conclude that recombination is important, but not more so than mutation is, in generating new combinations of PBR motifs. In addition, our study suggests that past recombination events which affected the present-day lineages seems evenly distributed along the sequence. In fact, the marginal maximum likelihood estimate of ρ for each pair of sites and the composite likelihood are very similar in all comparisons (the difference being always smaller than the conventional threshold of 2). This result is compatible with those obtained in other species (Richman *et al*. 2001). It also seems reasonable considering that PBR codons are scattered throughout exon 2 and diffused recombination (as estimated from the observed lineages which are probably old enough to have passed the scrutiny of natural selection) would create new PBR variants more efficiently.

Long-term balancing selection at MHC

As observed for many other species (Bernatchez & Landry 2003; Sommer 2005; Piertney & Oliver 2006), the molecular evolution of chamois MHC variation was clearly affected by balancing selection, that is, positive selection with long retention of selected alleles. Various sources of evidence converge on this hypothesis. In this section, we discuss the results of the most classical analysis based on long-term evolutionary patterns (d_n/d_s ratio and trans-species polymorphisms). The results related to the within- and between-population genetic variation indices will be discussed in the last section.

Each substitution in this region produced a different protein sequence in our sample (28 different proteins among

191 individuals), with an extremely high estimated d_n/d_s ratio. With the exception of a study on sockeye salmon by Miller *et al*. (2001), we are not aware of similar results in other species. It is not clear what selective and/or population mechanism can maintain different protein lineages and simultaneously eliminate DNA sequences with synonymous substitution for a long period of time. Richman *et al*. (2003a) hypothesize a combination of recombination, balancing selection and repeated bottlenecks, but we believe that the details of their model requires further clarification, possibly using a systematic simulation approach.

Polymorphisms that predate species divergence due to long retention of lineages, known as trans-species polymorphism, are considered as a clear evidence of balancing selection (Takahata 1990; Takahata & Nei 1990). We reported this pattern in the MHC sequences of the *Caprinae* family, and in particular within the genus *Rupicapra*. The Alpine and the Pyrenean chamois (*Rupicapra rupicapra* and *Rupicapra pyrenaica*, respectively) still share some MHC alleles, despite the fact that divergence of the species is estimated at about 1.6 million years (Lalueza-Fox *et al*. 2005), and despite the presence at the mtDNA cyt *b* sequences of strict monophyly with an average divergence between species about 2.5 times larger than the average divergence within species (B. Crestanello, personal observation). Clearly, some form of balancing selection is responsible for this pattern along the evolutionary history of the species. However, our data are compatible with overdominance, frequency-dependent selection, or a combination of the two (Takahata & Nei 1990; Piertney & Oliver 2006). The absence of significant deviation from the Hardy–Weinberg expectation in all populations we analysed can only exclude a strong overdominance effect.

The interplay between selection, demography and population structure

The long-term effects of balancing selection throughout the phylogenetic history of the chamois species are now clear. However, understanding if and to what extent diversity within and between populations preserves the signature of natural selection, or are rather affected mainly by demographic processes, requires additional analyses.

The results of the neutrality tests are interesting, but only partially informative. The Ewens–Watterson–Slatkin test seems to exclude that selection is currently playing an important role in changing allele frequencies within the populations we analysed. The sample configuration within each of the groups is as even (or uneven) as expected by the neutrality assumption, suggesting that positive selection at single alleles or genotypes has not affected population dynamics in the short term. On the other hand, (i) the sign and the values produced by Tajima's *D* and Fu's *F*-tests clearly suggest that neutrality should be excluded at least within the evolutionary time frame relevant for the accumulation of mutations, and (ii) the analysis of the geographical structure shows that the pattern of genetic variation between populations is compatible with the active maintenance by selection of the same set of MHC alleles in otherwise differentiated groups. We will consider these two points in turn.

First of all, taking recombination into account when computing the null distribution of Tajima's *D* and Fu's *Fs* can significantly affect the outcomes. Recombination tends to decrease the range of variation of Tajima's *D* under the null hypothesis (Wall 1999), which means that testing this hypothesis without considering recombination decreases the power of the test. Fu's *Fs*, on the other hand, is biased towards negative values if recombination is occurring, because every recombination event produces new alleles leaving the number of segregating sites unchanged. Therefore, this test can be either too permissive (if a selective sweep occurred or the population size increased in the past (Ramos-Onsins & Rozas 2002) or too conservative (if balancing selection is operating or the population size decreased in the past). Controlling for recombination led to an increase in the number of significant values of *D* and *Fs* from 6 to 14 in the ten chamois populations we analysed, and all these values were positive.

This result strongly supports the view that genetic variation is not neutral, and it also excludes the hypothesis that few MHC alleles diffused in these chamois groups as a results of a selective sweep or a recent population expansion. In fact under the latter circumstances, we would have obtained negative values in the of Tajima's *D* and Fu's *Fs* statistics. However, this does not directly tell us, whether balancing selection exclusively produced trans-species polymorphisms and an extreme d_n/d_s ratio especially in PBR region, or if it was also important in shaping the genetic variation pattern within groups. In fact, the same pattern of genetic variation can be simply related to a demographic event of population size contraction.

The demography of the chamois populations, here analysed using the mtDNA control region sequence, revealed a likely reduction in population size in the last 5000 to 30 000 years. Contrary to many other species which experienced a demographic contraction within Mediterranean refugia during the last glaciation, and a subsequent northern expansion when the temperature increased (Petit *et al*. 2003), the results we obtained support the view that the chamois followed a different demographic trajectory. Increased temperatures reduced the available territory for the chamois and confined them to mountain peaks, with the likely result of decreasing population size. Therefore, demographic analysis of mtDNA suggests that the Tajima's *D* and Fu's *Fs* test results for the MHC locus could be explained by a demographic effect only. Nonetheless, we believe that both demography and balancing selection influenced the pattern of genetic variation within and between chamois populations. We justify this view considering in detail patterns of geographical variation.

The average population divergence at the MHC locus, estimated using F_{ST}, was about 10 times smaller than that observed at the mtDNA sequences (5.2% vs. 56.6%, respectively). Can this difference be simply explained by the four times larger effective population size at MHC compared to the mtDNA? Or is an additional process (i.e. selection) shaping the MHC genetic structure? Assuming a simple infinite island model at mutation–drift equilibrium, constant population size, sex-unbiased migration, and no selection, the expected divergence at nuclear markers can be derived using the estimated divergence at mtDNA (Crochet 2000). Using this approach for the chamois (see for example Brito 2007), we obtain a predicted F_{ST} at the nuclear MHC locus equal to 24.6%, five times larger than the observed value. This approach, however, is clearly too simplistic, so we performed a simulation experiment assuming a more realistic demographic scenario. The results of this experiment indicate that the large difference between F_{ST} values at mtDNA and MHC cannot be generated by random drift only.

In principle, sex-biased migration rates, with males dispersing more than females, as observed in two game reserve populations (Loison *et al*. 1999), could explain the results of the simulation. However, this same pattern was not seen in an undisturbed population in the Gran Paradiso National Park (Lovari *et al*. 2006). Furthermore, even increasing the migration rate at the nuclear locus by a factor of 3, the distribution of the ratio between mitochondrial and nuclear F_{ST} in the simulations was smaller than the value observed in approximately 90% of replicates.

F_{ST} estimates based on MHC and neutral markers have been compared in different species. In most cases, the MHC population structure was either similar or higher than observed at other loci (Landry & Bernatchez 2001; Miller *et al*. 2001; Cohen 2002; Ekblom *et al*. 2007; Alcaide *et al*. 2008; see also Table 1 in Muirhead 2001), and these results were attributed to weak selection at MHC, local adaptation, or problems in the comparison between markers with different mutation models. In contrast, we observed a pattern similar to that of two Trinidadian guppy populations (van Oosterhout *et al*. 2006), with very little divergence at MHC compared to neutral markers. The same set of MHC alleles is maintained in different populations (see Fig. 4), otherwise isolated and highly subject to drift effects. This could be partially explained by increased effective migration rate at MHC, theoretically predicted under a model of symmetric overdominance (Schierup *et al*. 2000; Muirhead 2001). However, migration rates among chamois groups are probably very low in general, since this species can be found only at high altitudes and the valleys (especially in recent times due to urbanization) represent strong barriers to gene flow. In addition, our results seem to exclude strong overdominance effects. It is more likely that the geographical distribution of MHC alleles is shaped by similar pathogen communities across the ecologically homogenous area we considered, coupled with a balancing selection process acting within groups. This process is preventing both population differentiation and — as observed for example in the island fox (Aguilar *et al*. 2004) — the reduction of MHC genetic variation within groups during fragmentation and bottlenecks.

Conclusions

In this study, we showed that balancing selection shaped, probably through a process which implies limited allele frequencies dynamic in the short term, the major aspects of the genetic variation in the DRB exon 2 locus in the Alpine chamois. Alleles coding for different proteins are maintained along the long evolutionary history of the species, but also within geographically related, but genetically isolated, populations highly subjected to drift effects. In particular, the combined results from MHC and mtDNA sequences support a scenario whereby the last glaciation produced in the chamois populations a demographic decline and a fragmentation in small subunits. These groups, however, preserved since then much of their variation and continuity at the MHC locus through a balancing selection process which favoured the same set of alleles in different populations. It is worth noting that, besides its evolutionary significance, balancing selection should be also considered an important mechanism to counterbalance the loss of diversity at the very important MHC loci. The open question is how fast this mechanism can operate in species, such as the chamois, where human activities constantly introduce or enhance barriers to gene flow and interfere with the natural population dynamics through hunting and translocations.

Acknowledgements

This study was supported by the Dolomiti Bellunesi National Park, the University of Ferrara, the University of Torino, the Italian Ministry of Universities (MIUR PRIN 2005) and the Centre of Alpine Ecology (Trento). The Corpo di Polizia Provinciale di Belluno contributed to the sampling. We also thank R. Permunian, P. Zandegiacomo, G. Fiori L. Grones, and E. Vettorazzo for their help.

References

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences, USA*, **9**, 3490–3394.

Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at the positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, **17**, 2652–2665.

Alvarez-Busto J, García-Etxebarria K, Herrero J, Garin I, Jugo BM (2007) Diversity and evolution of the Mhc-DRB1 gene in the two endemic Iberian subspecies of Pyrenean chamois, *Rupicapra pyrenaica*. *Heredity*, **99**, 406–413.

Amills M, Jiménez N, Jordana J *et al*. (2004) Low diversity in the major histocompatibility complex class II DRB1 gene of the Spanish ibex, *Capra pyrenaica*. *Heredity*, **93**, 266–272.

Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics and Development*, **11**, 635–641.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179–224.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G (2006) Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1111–1116.

Bos DH, Turner SM, Dewoody JA (2007) Haplotype inference from diploid sequence data: evaluating performance using non-neutral MHC sequences. *Hereditas*, **144**, 228–234.

Brito PH (2007) Contrasting patterns of mitochondrial and microsatellite genetic structure among Western European populations of tawny owls (*Strix aluco*). *Molecular Ecology*, **16**, 3423–3437.

Brown JH, Jardetzky TS, Gorga JC *et al*. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **64**, 33–39.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Cohen S (2002) Strong positive selection and habitat specific substitution patterns in Mhc from an estuarine fish under intense pollution stress. *Molecular Biology and Evolution*, **19**, 1870–1880.

Crochet PA (2000) Genetic structure of avian populations allozymes revisited. *Molecular Ecology*, **9**, 1463–1469.

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

Douzery E, Randi E (1997) The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Molecular Biology and Evolution*, **14**, 1154–1166.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

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

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

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

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

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

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences (with discussion). *Statistical Sciences*, **7**, 457–511.

Gutierrez-Espeleta GA, Kalinowski ST, Boyce WM, Hedrick PW (2000) Genetic variation and population structure in desert bighorn sheep: implications for conservation. *Conservation Genetics*, **1**, 3–15.

Gyllensten UB, Sundvall M, Erlich HA (1991) Allelic diversity is generated by intraexon sequence exchange at the DRB1 locus of primates. *Proceedings of the National Academy of Sciences, USA*, **88**, 3686–3690.

Hartl DL, Clark AG (2006) *Principles of Population Genetics*, 4th edn. Sinauer & Associates Inc, Sunderland, Massachusetts.

Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. *American Naturalist*, **143**, 945–964.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629–636.

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

Hedrick PW, Kim T (2000) Genetics of complex polymorphisms: parasites and maintenance of MHC variation. In: *Evolutionary Genetics: From Molecules to Morphology* (eds Singh R, Krimbas C), pp. 204–234. Cambridge University Press, Cambridge, UK.

Hedrick PW, Parker KM, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex variation in the Arabian oryx. *Evolution*, **54**, 2145–2151.

Hess CM, Wang Z, Edwards SV (2007) Evolutionary genetics of *Carpodacus mexicanus*, a recently colonized host of a bacterial pathogen, *Mycoplasma gallisepticum*. *Genetica*, **129**, 217–225.

Hosomichi K, Shiina T, Suzuki S *et al*. (2006) The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics*, **7**, 322.

Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics*, **159**, 1805–1817.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.

Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. *Heredity*, **81**, 239–245.

Klein J (1980) Generation of diversity at MHC loci: implications for T-cell receptor repertoires. In: *Immunology 80* (eds Fougereau M, Dausset J), pp. 239–253. Academic Press, London.

Klein J (1986) *Natural History of the Major Histocompatibility Complex*, 1st edn. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore.

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

Kreitman M (2000) Methods to detect selection in populations with applications to the human. *Annual Review of Genomics and Human Genetics*, **1**, 539–559.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lalueza-Fox C, Castresana J, Sampietro L, Marquès-Bonet T, Alcover JA, Bertranpetit J (2005) Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. *BMC Evolutionary Biology*, **5**, 70.

Landry C, Bernatchez L (2001) Comparative analysis of population structure across environments and geographical scales at major

histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **10**, 2525–2539.

Laval G, Excoffier L (2004) SimCoal 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**, 2213–2233.

Loison A, Jullien J-M, Menaut P (1999) Subpopulation structure and dispersal in two populations of chamois. *Journal of Mammalogy*, **80**, 620–632.

Lovari S, Sacconi F, Trivellini G (2006) Do alternative strategies of space use occur in male Alpine chamois? *Ethology Ecology and Evolution*, **3**, 221–231.

Mannen H, Nagata Y, Tsuji S (2001) Mitochondrial DNA reveal that domestic goat (*Capra hircus*) are genetically affected by two subspecies of Bezoar (*Capra aegagurus*). *Biochemical Genetics*, **39**, 145–154.

Marchini J, Cutler D, Patterson N *et al*. (2006) International HapMap Consortium. A comparison of phasing algorithms for trios and unrelated individuals. *American Journal of Human Genetics*, **78**, 437–450.

McVean G, Awadalla P, Fearnhead PA (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231–1241.

Meyer D, Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1–26.

Meyer D, Single RM, Mack SJ, Erlich HA, Thomson G (2006) Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. *Genetics*, **173**, 2121–2142.

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

Muirhead CA (2001) Consequences of population structure on genes under balancing selection. *Evolution*, **55**, 1532–1541.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nevo E, Beiles A (1992) Selection for class II MHC heterozygosity by parasites in subterranean mole rats. *Experientia*, **48**, 512–515.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nylander JA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

van Oosterhout C, Joyce DA, Cummings SM *et al*. (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*, **60**, 2562–2574.

Otto SP (2000) Detecting the form of selection from DNA sequence data. *Trends in Genetics*, **16**, 526–529.

Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. *Science*, **272**, 67–74.

Penn J, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist*, **153**, 145–164.

Pérez T, Albornoz J, Domínguez A (2002) Phylogeography of chamois (*Rupicapra* spp.) inferred from microsatellites. *Molecular Phylogenetics and Evolution*, **25**, 524–534.

Petit RJ, Jouzel J, Raynaud D *et al*. (1999) Climate and atmospheric history of the past 420 000 years from the Vostok ice core, Antarctica. *Nature*, **399**, 429–436.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Ploegh H, Watts C (1998) Antigen recognition. *Current Opinion in Immunology*, **10**, 57–58.

Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, **5**, 181–186.

Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics*, **160**, 1179–1189.

Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, **155**, 1429–1437.

Rambaut A, Drummond AJ (2004) *Tracer*. University of Oxford, Oxford, UK.

Ramos-Onsins SE, Mithell-Olds T (2007) MLCOALSIM: multilocus coalescent simulations. *Evolutionary Bioinformatics*, **2**, 41–44.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, **43**, 304–311.

Richman AD, Herrera LG, Nash D (2001) MHC class II beta sequence diversity in the deer mouse (*Peromyscus maniculatus*): implications for models of balancing selection. *Molecular Ecology*, **10**, 2765–2773.

Richman AD, Herrera LG, Nash D (2003a) Evolution of MHC class II E beta diversity within the genus *Peromyscus*. *Genetics*, **164**, 289–297.

Richman AD, Herrera LG, Nash D, Schierup MH (2003b) Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. *Genetical Research*, **82**, 89–99.

Robinson J, Waller MJ, Parham P *et al*. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Research*, **31**, 311–314.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sanchez-Mazas A (2007) An apportionment of human HLA diversity. *Tissue Antigens*, **69**(S1), 198–202.

Satta Y, Takahata N, Schönbach C, Gutknecht J, Klein J (1991) Calibrating evolutionary rates at major histocompatibility complex loci. In: *Molecular Evolution of the Major Histocompatibility Complex* (eds Klein J, Klein D), pp. 51–62. Springer-Verlag, Heidelberg, Germany.

Schaschl H, Goodman SJ, Suchentrunk F (2004) Sequence analysis of the MHC class II DRB alleles in Alpine chamois (*Rupicapra r. rupicapra*). *Developmental and Comparative Immunology*, **8**, 265–277.

Schaschl H, Suchentrunk F, Hammer S, Goodman SJ (2005) Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.). *Immunogenetics*, **57**, 108–115.

Schaschl H, Wandeler P, Suchentrunk F, Obexer-Ruff G, Goodman SJ (2006) Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. *Heredity*, **97**, 427–437.

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics*, **156**, 879–891.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research*, **76**, 51–62.

Schierup MH, Mikkelsen AM, Hein J (2001) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics*, **159**, 1833–1844.

Schwensow N, Fietz J, Dausmann K, Sommer S (2007a) MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology, doi* 10.1007/s10682-007-9186-4.

Schwensow N, Fietz J, Dausmann KH, Sommer S (2007b) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity*, **99**, 265–277.

Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research*, **64**, 71–74.

Slatkin M (1996) A correction to the exact test based on the Ewens sampling distribution. *Genetical Research*, **68**, 259–260.

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

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceedings of the National Academy of Sciences, USA*, **87**, 2419–2423.

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967–978.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Trowsdale J (1993) Genomic structure and function in the MHC. *Trends in Genetics*, **9**, 117–122.

Wakeland EK, Boehme S, She JX *et al*. (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. *Immunological Research*, **9**, 115–122.

Wall JD (1999) Recombination and the power of statistical tests of neutrality. *Genetical Research*, **74**, 65–79.

van der Walt JM, Nel LH, Hoelzel AR (2001) Characterization of major histocompatibility complex DRB diversity in the endemic South African antelope *Damaliscus pygargus*: a comparison in two subspecies with different demographic histories. *Molecular Ecology*, **10**, 1679–1688.

Watterson GA (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405–417.

Zeng K, Mano S, Shi S, Wu CI (2007) Comparisons of site- and haplotype-frequency methods for detecting positive selection. *Molecular Biology and Evolution*, **24**, 1562–1574.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Alignment of the 28 different MHC class II DRB exon 2 amino acid sequences found in this study. Dots indicate identity with the Ruru-DRB01 sequence; stars indicate codons belonging to the putative PBR region, following the structure of the human orthologue (Brown *et al*. 1993).

Table S1 List of alleles (from GenBank) used in the phylogenetic analysis

Table S2 Number of MHC DRB exon 2 alleles in the 10 populations

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the *Molecular Ecology* Central Office.