

ORIGINAL ARTICLE

Selection and recombination drive the evolution of MHC class II *DRB* diversity in ungulates

H Schaschl¹, P Wandeler², F Suchentrunk³, G Obexer-Ruff⁴ and SJ Goodman⁵

¹Konrad Lorenz Institute for Ethology, Vienna, Austria; ²Zoological Museum, University of Zurich-Irchel, Zurich, Switzerland;

³Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Vienna, Austria; ⁴Institute for Genetics, Nutrition and Husbandry of Companion Animals, Berne, Switzerland and ⁵Institute of Integrative and Comparative Biology, University of Leeds, Leeds, UK

Major histocompatibility complex (MHC) antigen-presenting genes are the most variable loci in vertebrate genomes. Host–parasite co-evolution is assumed to maintain the excessive polymorphism in the MHC loci. However, the molecular mechanisms underlying the striking diversity in the MHC remain contentious. The extent to which recombination contributes to the diversity at MHC loci in natural populations is still controversial, and there have been only few comparative studies that make quantitative estimates of recombination rates. In this study, we performed a comparative analysis for 15 different ungulates species to estimate the population recombination rate, and to quantify levels of selection. As expected for all species, we observed signatures of strong positive selection, and identified individual residues experiencing selection that were congruent with those constituting the peptide-binding region of

the human *DRB* gene. However, in addition for each species, we also observed recombination rates that were significantly different from zero on the basis of likelihood-permutation tests, and in other non-quantitative analyses. Patterns of synonymous and non-synonymous sequence diversity were consistent with differing demographic histories between species, but recent simulation studies by other authors suggest inference of selection and recombination is likely to be robust to such deviations from standard models. If high rates of recombination are common in MHC genes of other taxa, re-evaluation of many inference-based phylogenetic analyses of MHC loci, such as estimates of the divergence time of alleles and trans-specific polymorphism, may be required.

Heredity (2006) **97**, 427–437. doi:10.1038/sj.hdy.6800892; published online 30 August 2006

Keywords: MHC class II; recombination; selection; mutation; allelic polymorphism; ungulates

Introduction

Major histocompatibility complex (MHC) antigen-presenting genes (classes I and II) are the most variable loci in vertebrate genomes. MHC class I molecules are proteins that are expressed on the surface of all nucleated cells and present peptides to CD8⁺ T cells. Class II MHC molecules, which consist of non-covalently associated α and β chains, are expressed only on antigen-presenting cells such as B cells, dendritic cells and macrophages. The α_1 and β_1 domains, which are encoded by separate genes, form the peptide-binding region (PBR), in which peptides are bound and recognized by CD4⁺ helper T cells. Three distinct isoforms of class II MHC molecules are expressed in human. They are denoted HLA-*DR*, -*DQ* and -*DP*. The *DRB1* locus is the most polymorphic class II locus known in humans with approximately 360 alleles described (Robinson *et al.*, 2003). The extreme level of polymorphism observed for these genes is thought to arise from the need to recognize a diverse

repertoire of pathogens that exhibit antagonistic co-evolution with the host to evade immune recognition (Doherty and Zinkernagel, 1975), and from potential roles in mate recognition (or inbreeding avoidance) for some species (Jordan and Bruford, 1998; Penn and Potts, 1999; Reusch *et al.*, 2001). The pathogen-driven selection hypothesis is supported by the characteristic patterns of molecular evolution exhibited by MHC class I and class II genes in most species, including significant excesses of non-synonymous mutations over synonymous mutations for residues involved in peptide binding, evidence of balancing selection (Hughes and Nei, 1988) and a ‘trans-species’ mode of evolution, where many alleles appear to be older than the species in which they are found (Figueroa *et al.*, 1988; Klein *et al.*, 1993).

While MHC variation is getting increasing attention for a growing number of species (Bernatchez and Landry, 2003), there is still debate over the relative importance of different mechanisms for generating the large allelic diversity seen at these MHC loci. From the first MHC studies at a molecular level, most attention has focused on pathogen-mediated selection as the primary mechanism driving the evolution of MHC diversity. Studies focusing on the pattern of molecular evolution show clear evidence of the imprint of selection by pathogens (Hughes and Yeager, 1998; Hughes, 2002). In contrast, for natural or seminatural population studies, direct empiri-

Correspondence: Dr H Schaschl, Konrad Lorenz Institute for Ethology, Austrian Academy of Sciences, Vienna, Savoyenstrasse 1a, A-1160 Vienna, Austria.

E-mails: H.Schaschl@gmx.net or s.j.Goodman@leeds.ac.uk

Received 8 June 2006; revised 29 June 2006; accepted 3 July 2006; published online 30 August 2006

cal demonstration of the differential fitness of alternate MHC genotypes in response to pathogens has been more challenging, with few studies possessing the necessary supporting data and statistical power to detect subtle effects (Paterson *et al.*, 1998; Lohm *et al.*, 2002; Penn *et al.*, 2002; Wegner *et al.*, 2003).

Despite the importance of recombination in population genetics, remarkably little is known regarding the extent and rate of recombination, particularly in wild populations. Recombination rates are known to change across the human genome (McVean *et al.*, 2004), but also between very closely related species (Ptak *et al.*, 2004). Recent work has suggested that intragenic recombination may also be an important mechanism for the generation of allelic diversity at MHC class I and class II loci (Gyllensten *et al.*, 1991; She *et al.*, 1991; Andersson and Mikko, 1995; Satta, 1997; Bergstrom *et al.*, 1998; Takahata and Satta, 1998; Mikko *et al.*, 1999; Richman *et al.*, 2003a; Schaschl *et al.*, 2005). While mutations generate novel variation at individual sites, recombination can produce multiple substitutions per event with the potential for greater change in the antigen-binding properties of MHC molecules. Therefore, novel molecules generated by recombination may have a greater selective advantage (She *et al.*, 1991; Andersson and Mikko, 1995). Assessing the relative importance of selection, recombination and mutation in generating the diversity seen at MHC genes relies on being able to make quantitative estimates of these parameters, but robust quantitative population genetic estimators for the amount of recombination have been lacking until recently. Therefore, previous studies of recombination in the MHC have often relied on non-parametric methods that can detect recombination in a sample of sequences, but without estimating the rate at which it occurs (e.g., Andersson and Mikko, 1995). Although previous quantitative analysis of recombination has been impeded by lack of a strong theoretical framework, new developments in this area are now permitting progress (Hudson, 2001; McVean *et al.*, 2002; Stumpf and McVean, 2003). The approach of McVean *et al.* (2002) is a composite likelihood method (Hudson, 2001), based on coalescent theory, employing a finite sites mutation model (e.g., allow recurrent mutations) to estimate the population recombination rate ($4N_e r$). This is combined with a powerful likelihood-permutation test to determine if observed values of $4N_e r$ are significantly different from zero. Both the estimator and significance test are robust to some misspecification of the model of sequence evolution (Richman *et al.*, 2003b). This new quantitative method has so far been applied to a limited number of taxa for the analysis of MHC sequences. For example, Richman *et al.* (2003a,b) found evidence of extensive recombination at the MHC class II *DRB* gene in deer mouse (*Peromyscus*) taxa; the population recombination rate is significantly different from zero and greatly exceeds the population mutation rate. However, whether this pattern observed in *Peromyscus* species is representative of other taxa, and the importance of recombination as a general mechanism involved in the evolution of MHC diversity remains to be demonstrated. More comparative surveys of multiple species are an important contribution to this debate as population demography also plays a key role, and the apparent importance of selection and recombination may change after fluctuations in population size (Richman *et al.*, 2003a). Quanti-

fying the relative importance of recombination in MHC genes has important implications for our understanding of the evolution of these loci. Failure to account for recombination may significantly influence population genetic and phylogenetic inferences (Schierup and Hein, 2000; Schierup *et al.*, 2001), affecting interpretations of MHC allelic genealogies, and estimates of the strength and mode of selection.

In this paper, we present a comparative analysis of MHC class II *DRB* diversity in 15 ungulate species and, in particular, we test whether recombination has played a significant role in the evolution of MHC class II *DRB* diversity.

Materials and methods

Samples and DNA isolation

Liver samples were obtained from 27 Alpine ibex (*Capra ibex*) from one population in Switzerland (Albris, Kanton Graubunden, Switzerland), seven Spanish ibex (*Capra pyrenaica*) from the populations Tortosa-Beceite and Parque Natural Sierra de las Nieves – Ronda (Spain) and 18 Himalayan tahr (*Hemitragus jemlahicus*) from New Zealand. These animals were taken during hunts for population management purposes by licensed managing authorities, and were not killed specifically for this study. A blood sample of one tahr individual was collected at Zoo Schönbrunn (Vienna, Austria). Genomic DNA was extracted from frozen liver using standard phenol/chloroform-isoamyl alcohol extraction or with the Qiagen DNA extraction kit. The tahr blood sample was used to isolate mRNA, and this was converted into cDNA using the Titan One Tube RT-PCR-System from Roche Diagnostic GmbH, Vienna, Austria.

PCR, cloning and sequencing

MHC class II *DRB* exon 2 (β_1 domain) sequences were obtained by PCR using locus-specific primers in a seminested PCR approach as described in Schaschl *et al.* (2004). PCR products were cloned into the TA vector pCR2.1 (Invitrogen Ltd., Karlsruhe, Germany), and then transformed in *Escherichia coli*-competent cells. From each individual, between five and seven positive (blue/white selection) clones were picked and grown overnight in 3 ml of LB-Medium. Plasmids were isolated by the QIAprep Spin Miniprep Kit (Qiagen) and subsequently checked for the correct inserts with *EcoRI* digestion or by running a second PCR with M13 primers. Those clones, which contained the correct inserts, were sequenced bi-directionally with the Applied Biosystems BigDye Terminator v2.0 Cycle Sequencing Kit, using M13 universal and M13 reverse primers, on the ABI PRISM 377 DNA Sequencer. Samples with poor cloning efficiency and sequences produced by PCR recombination were discarded. The Alpine ibex samples were directly sequenced, because we expected low level of heterozygosity. However, from five random samples, the PCR products were cloned and sequenced as described above. Sequence data were prepared for further analysis with BioEdit (Hall, 1999).

Sequence analysis and detecting positive selection

Sequences were aligned using the program ClustalX (Thompson *et al.*, 1997). The selective pressure at the

amino-acid level is measured by the non-synonymous/synonymous rate ratio ($\omega = d_N/d_S$). A ω ratio greater than 1 ($\omega = d_N/d_S > 1$) indicate positive selection (e.g., Yang and Bielawski, 2000). We used the programme CODEML of the PAML 3.14 package (Yang, 1997) to test for the presence of codon sites affected by positive selection and to identify those sites by a Bayesian approach implemented in CODEML. The models considered in this study were M7 (beta) and M8 (beta& ω) (Yang *et al.*, 2000). While recombination can potentially generate false positives in the detection of positive selection, these two models are much more robust against the occurrence of recombination than the other models implemented in CODEML (Anisimova *et al.*, 2003). Bayes' prediction of sites under positive selection also appears to be robust to recombination effects (Anisimova *et al.*, 2003). Under the model M7 (beta), the ω ratio varies according to the beta distribution (with parameters p and q) and does not allow for positive selected sites ($0 < \omega < 1$), and thus serves as the null model for comparison with model M8 (beta& ω). Model M8 adds an additional site class to the beta model to account for sites under positive selection ($\omega > 1$). For each species, the M7 and M8 models can be compared using a likelihood-ratio test (Nielsen and Yang 1998). The likelihood-ratio test statistics calculates twice the log-likelihood difference (2) compared with a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the two models. To provide phylogenetic information required for the analysis (Yang, 1997), the best tree for DRB sequences in each species was identified by maximum likelihood under the one-ratio model (M0) in CODEML. The average number of non-synonymous substitutions per non-synonymous site (d_N) and synonymous substitutions per synonymous site (d_S) was estimated with the MEGA 2.1 program (Kumar *et al.*, 2001), using the distance of Nei and Gojobori (1986) with the Jukes and Cantor (1969) correction for multiple substitutions. Standard errors were calculated by 1000 bootstrapping replicates.

Estimation of population selection, mutation and recombination rates

Since we lack information about effective population size for each species, estimates of the rates of selection, recombination and mutation have to be limited to the composite population parameters S , ρ and θ , respectively, which are the rates of selection (s), recombination per base (r) and mutation per base (u), scaled by effective population size. In trying to estimate these parameters simultaneously, however, an inherent problem arises, in that the model framework for estimating S assumes an absence of recombination, and estimates for ρ are made under the assumption of no selection. Yet, in the course of this work, we show that both these processes influence the sequences being studied. Current methodologies offer no easy fix to these important violations of underlying assumptions, but attempts have been made to try and quantify the bias likely to be introduced (Satta, 1997; Fearnhead and Donnelly, 2001; Richman *et al.*, 2003a; Smith and Fearnhead, 2005). The main objective of this study is to try and assess the relative magnitudes of selection, recombination and mutation. The existing simulation studies suggest that while absolute values of

parameters are likely to be compromised by violation of model assumptions, relative values can be accepted with more confidence. Therefore, we proceed by acknowledging the likely limitations on this work, starting by outlining key model assumptions in the following methods, and then provide a discussion of the limitations of current statistical methods in relation to our results in the final part of the paper.

Estimation of S

Under a model of symmetric balancing selection (Kimura and Crow, 1964; Takahata and Nei, 1990), population selection can be defined as $S = 2Ns$, where N is the number of breeding individuals (effective population size) and s is the selective disadvantage of homozygotes. In this model, there are a fixed number of alleles (n) through time, allele frequencies are approximately evenly distributed and there is allelic turnover, so that when a new allele arises, it either replaces its parental allele in the genealogy or one of the other allelic lineages becomes extinct and is replaced. Here 'alleles' refers to functional alleles, and each new allele arises through a single non-synonymous mutation in the PBR according to an infinite allele model. Additional assumptions are finite sites, no recombination and there is no inherent correction for multiple mutations at the same site. Takahata *et al.* (1992) show that under these conditions, S is proportional to the number of functional alleles and the selection intensity, which can be estimated as:

$$S = \frac{n^2\gamma}{\sqrt{2}}$$

where n is the number of functional alleles in the population and γ is the ratio of the number of non-synonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site to ($(K_n/L_n)/(K_s/L_s)$), where n is approximated from K_n , which is the average pairwise number of non-synonymous substitutions in the PBR among sequences (Takahata *et al.*, 1992; Richman *et al.*, 2003a). K_n , L_n , K_s and L_s were estimated from the data, using the Jukes and Cantor (1969), within the Mega 2.1 program (Kumar *et al.*, 2001). Takahata *et al.* (1992) advocate that if correction for multiple substitutions is to be made, that should be done by selecting the data so as to include only 'young' alleles in which sites have not had chance to experience saturation. However, since recombination is likely to make assessing the apparent age of alleles problematic, to ensure consistent treatment of data across all species (and with the data used to estimate ρ), while maintaining sample sizes for species with low numbers of sequences, we chose to include all sequences for each species rather than to attempt a correction.

Estimation of ρ and θ

The LDhat program was used to estimate rates of population mutation (from Waterson's θ , where $\theta = 4N_e u$) and population recombination (ρ , where $\rho = 4N_e r$) (see McVean *et al.*, 2002). This program estimates population recombination from a set of aligned sequences using the composite-likelihood method (Hudson, 2001) within a coalescent framework. For the estimation of θ and ρ , the analysis was restricted to bi-allelic sites where the frequency for the alternate allele was ≥ 0.1 . This

approach was taken because the method does not consider among site rate variation and excluding rare allelic variants can increase the power of the analysis (McVean *et al.*, 2002). We used the likelihood permutation-based approach to test significance against the null hypothesis of no recombination (i.e., $4N_e r = 0$). The LDhat program recovers accurate estimates of relative recombination rate even when assumptions of the standard model are violated, such as through the action of selection (Richman *et al.*, 2003a; Smith and Fearnhead, 2005). However, under these circumstances, technically it may no longer be a true estimator of the population recombination rate ($4N_e r$). We also calculated the ratio ρ/θ as an estimate of the relative amount of recombination compared to point mutations, which again is robust to deviations from the underlying coalescent model (Fearnhead and Donnelly, 2001), and the corresponding ratio S/θ . Meunier and Eyre-Walker's (2001) G4 and Awadalla *et al.*'s (1999) r^2 statistics, which assess the correlation of linkage disequilibrium among pairs of polymorphic sites with the distance between them, were calculated as alternate indicators of recombination. In these cases, recombination is inferred if there is a significant decay of linkage disequilibrium with distance. Finally, we used the program DnaSP 4 (Rozas *et al.*, 2003) to calculate the minimum number of inferred recombination events (R_M) (Hudson and Kaplan, 1985) for each species, which can be interpreted as a minimal number of different positions at which recombination has occurred.

Sequence nomenclature and sequences from GenBank
In accordance with the proposed nomenclature for MHC in non-human species (Klein *et al.*, 1990), we designated the exon 2 alleles *Caib-DRB* for Alpine ibex (*Capra ibex*), *Capy-DRB* for Spanish ibex (*Capra pyrenaica*) and *Heje-DRB* for Himalayan tahr (*H. jemlahicus*) with serial numbers attached. Further, MHC class II *DRB* exon 2 sequence data from different ungulate species were taken from GenBank with following Accessions numbers: Alpine chamois (*Rupicapra r. rupicapra*), Accession nos.: AY368437–AY368455; Pyrenean chamois (*Rupicapra pyrenaica*), Accession nos.: AY212149–AY212157; Bighorn-sheep (*Ovis canadensis*), Accession nos.: AF324840–AF324861; wild goat (*Capra aegagrus*), Accession nos.: U00183–U00202; sheep (*O. aries*), Accession nos.: U00204–U00233; domestic cattle (*Bos taurus*), Accession nos.: U00124–U00144; water buffalo (*Bubalus bubalus*), Accession nos.: AF385473–AF385480; African buffalo

(*Syncerus caffer*), Accession nos.: AF059233–AF059241 (the sequence AF059236 was later excluded from the analyses due to potential sequencing artifacts); South African antelope, consistent of the two subspecies bontebok (*Damaliscus pygargus pygargus*) and blesbok (*Damaliscus p. phillipsi*), Accession nos.: AJ302736–AJ302762; white-tailed deer (*Odocoileus virginianus*), Accession nos.: AF082161–AF082175 and AF407169–AF407171; reindeer (*Rangifer tarandus*), Accession nos.: AF012716–AF012724; and moose (*Alces alces*), Accession nos.: X83278–X83286. Although much *DRB* sequence data is available for other ungulate species, we restrict this analysis to data that can be considered unambiguously derived from single loci and potential functionally, in order to avoid bias that would be introduced due to the inadvertent inclusion of multiple loci, as might be possible when non-locus-specific primers are used to amplify MHC class II loci from genomic DNA.

Results

Sequence analysis of novel exon 2 sequences

The novel *DRB* exon 2 (β_1 domain) sequences obtained for this study were 284 bp in length including primer sequences. The nucleotide sequences excluding primers (236 bp) of the exon 2 from each species have been deposited in GenBank (Accession numbers: AY706312 for Alpine ibex and AY706313–AY706317 for Himalayan tahr). No more than two sequences were detected for any individual in the analyzed samples, and we are confident that products were amplified from a single MHC class II *DRB* locus. We obtained four exon 2 alleles from the Himalayan tahr from New Zealand samples and a fifth distinct allele (as a homozygous genotype) from the zoo sample (Figure 1). Among these alleles, 15.7% of 236 nucleotide sites and 29.5% of 78-amino-acid sites were polymorphic. Two exon 2 alleles were obtained from the Spanish ibex samples (Figure 1). These two alleles corresponded to two published exon 2 sequences on GenBank, named *Capy-DRB1-1* and *Capy-DRB1-2* (Accession nos.: AF461692 and AF461693; Amills *et al.*, 2004). An additional four Spanish ibex alleles (*Capy-DRB1-3*, *Capy-DRB1-4*, *Capy-DRB1-5* and *Capy-DRB1-6*) have previously been identified (GenBank Accession nos.: AF461694–AF461696, and AY351788; Amills *et al.*, 2004). These alleles were included in all subsequent analyses. For these exon 2 sequences 22.9% of 236 nucleotide sites

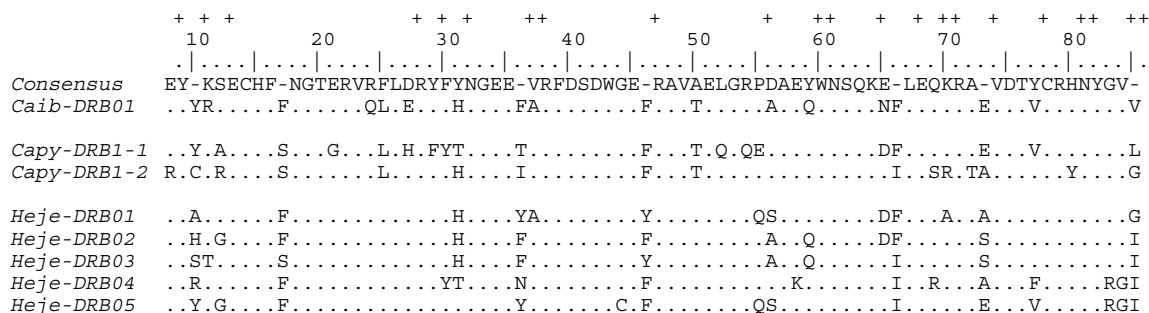


Figure 1 Alignment of the putative amino-acid sequences for MHC class II *DRB* exon 2 from Alpine ibex (*Caib-DRB*), Spanish ibex (*Capy-DRB*) and from Himalayan tahr (*Heje-DRB*). Dots indicate identity in the amino-acid sequence to the sequence of the *Consensus sequence* and a cross indicates codons involved in the PBR in humans (Brown *et al.*, 1993).

and 38.5% of 78-amino-acid sites were polymorphic. From the 27 Alpine ibex individuals, we obtained only a single exon 2 sequence (Figure 1). The single Alpine ibex sequence was in all subsequently analyses excluded.

Detecting positive selection at sites using maximum-likelihood analysis

Likelihood-ratio tests comparing the two models M7–M8 demonstrated that M8 (the model that accounts for sites under positive selection ($\omega > 1$)), fitted the data from all species significantly ($P < 0.001$), better than the model M7 (that does not allow for positive selection ($0 < \omega < 1$))

(Table 1). In all species, the ω ratio is over 1, indicating positive selection in the MHC class II DRB sequences. Sites identified as being under positive selection in the Bayesian analysis are listed in Table 2. Those sites that were identified as positively selected sites (at the 95% confidence level) are mostly in accordance with the human PBR sites (HLA-DRB1 gene) (Brown *et al.*, 1993). The identification of sites under positive selection is remarkably consistent across species, with many sites identified in most species (e.g., site 11 is identified in all except reindeer), and with majority identified in at least two species. The average number of synonymous substitutions per synonymous site (d_S) and

Table 1 Likelihood ratio test comparing the models M7–M8 for evidence of positive selection and model parameters estimates (ω is the selection parameter and p_n is the proportion of sites that fall into ω_n site class)

| Group | Species | Models estimates of parameters | | | | d_S | d_N |
|----------|-------------------------------|--------------------------------|---------|----------------------------|--|---------------|---------------|
| | | $2\Delta\ell$ | P-value | M7 (beta) | M8 (beta& ω) | | |
| Caprinae | Spanish ibex | 22.96 | <0.001 | $p = 0.016$ $q = 0.021$ | $p_0 = 0.848$ ($p_1 = 0.152$) $p = 0.019$ $q = 0.024$ $\omega = 8.81$ | 0.062 (0.019) | 0.142 (0.028) |
| | Himalayan tahr | 19.46 | <0.001 | $p = 0.005$ $q = 0.007$ | $p_0 = 0.943$ ($p_1 = 0.057$) $p = 0.005$ $q = 0.005$ $\omega = 17.46$ | 0.033 (0.015) | 0.112 (0.023) |
| | Alpine chamois ^a | 51.34 | <0.001 | $p = 0.016$ $q = 0.032$ | $p_0 = 0.869$ ($p_1 = 0.131$) $p = 2.490$ $q = 0.296$ $\omega = 15.23$ | 0.006 (0.003) | 0.061 (0.014) |
| | Pyrenean chamois ^a | 43.26 | <0.001 | $p = 0.005$ $q = 0.008$ | $p_0 = 0.771$ ($p_1 = 0.229$) $p = 0.117$ $q = 2.032$ $\omega = 20.49$ | 0.008 (0.005) | 0.066 (0.016) |
| | Wild goat | 67.62 | <0.001 | $p = 0.106$ $q = 0.177$ | $p_0 = 0.895$ ($p_1 = 0.105$) $p = 0.018$ $q = 0.024$ $\omega = 8.50$ | 0.055 (0.016) | 0.114 (0.022) |
| | Bighorn-sheep | 68.86 | <0.001 | $p = 0.016$ $q = 0.020$ | $p_0 = 0.937$ ($p_1 = 0.063$) $p = 0.005$ $q = 0.005$ $\omega = 15.67$ | 0.032 (0.012) | 0.115 (0.020) |
| | Sheep | 74.48 | <0.001 | $p = 0.016$ $q = 0.021$ | $p_0 = 0.929$ ($p_1 = 0.070$) $p = 0.014$ $q = 0.046$ $\omega = 7.46$ | 0.037 (0.011) | 0.096 (0.019) |
| Bovinae | Domestic cattle | 55.54 | <0.001 | $p = 0.015$ $q = 0.029$ | $p_0 = 0.823$ ($p_1 = 0.177$) $p = 0.005$ $q = 0.012$ $\omega = 5.72$ | 0.041 (0.013) | 0.114 (0.022) |
| | African buffalo | 15.40 | <0.001 | $p = 0.015$ $q = 0.019$ | $p_0 = 0.775$ ($p_1 = 0.225$) $p = 0.009$ $q = 0.034$ $\omega = 4.33$ | 0.079 (0.023) | 0.100 (0.021) |
| | Water buffalo | 15.14 | <0.001 | $p = 0.017$ $q = 0.022$ | $p_0 = 0.911$ ($p_1 = 0.089$) $p = 0.005$ $q = 0.008$ $\omega = 4.93$ | 0.106 (0.024) | 0.141 (0.024) |

Table 1 Continued

| Group | Species | Models estimates of parameters | | | | d_S | d_N |
|--------------|------------------------|--------------------------------|---------|----------------------------|--|---------------|---------------|
| | | $2\Delta\ell$ | P-value | M7 (beta) | M8 (beta& ω) | | |
| Alcelaphinae | South African antelope | 37.18 | <0.001 | $p = 0.005$ $q = 0.012$ | $p_0 = 0.740$ ($p_1 = 0.259$) $p = 0.005$ $q = 0.020$ $\omega = 7.42$ | 0.018 (0.009) | 0.096 (0.021) |
| Cervidae | Reindeer | 17.14 | <0.001 | $p = 0.005$ $q = 0.008$ | $p_0 = 0.773$ ($p_1 = 0.227$) $p = 0.028$ $q = 0.127$ $\omega = 5.07$ | 0.044 (0.013) | 0.088 (0.019) |
| | White-tailed deer | 82.37 | <0.001 | $p = 0.108$ $q = 0.187$ | $p_0 = 0.874$ ($p_1 = 0.126$) $p = 0.019$ $q = 0.025$ $\omega = 7.79$ | 0.035 (0.015) | 0.140 (0.025) |
| | Moose | 43.26 | <0.001 | $p = 0.005$ $q = 0.048$ | $p_0 = 0.933$ ($p_1 = 0.067$) $p = 0.005$ $q = 99.00$ $\omega = 57.16$ | 0.008 (0.008) | 0.026 (0.011) |

Average number of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) estimated for the exon 2 (77 codons) (d_S - d_N values in parentheses are s.e.'s) for each species is also given. Twice the difference in likelihood ($2\Delta\ell$) between the models was compared to a χ^2 distribution with d.f. = 2 to assess the significance of tests. (P-values. No data were estimated for the single obtained DRB allele of Alpine ibex.)

^aAlpine and Pyrenean chamois data were taken from Schaschl *et al.* (2005).

Table 2 The plus symbol presents sites under positive selection (model M8, at the >95% confidence level) identified by a Bayesian method implemented in CODEML

| Group | Species | Positively selected sites | | | | | | | | | | | | | | | | | | | | | |
|--------------|------------------------|---------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|---|
| | | 10 | 11 | 13 | 30 | 31 | 32 | 37 | 38 | 47 | 56 | 57 | 60 | 67 | 70 | 71 | 74 | 77 | 78 | 86 | | | |
| Caprinae | Spanish ibex | | + | + | | | + | | | | | | + | | + | + | | | | + | + | | |
| | Himalayan tahr | | + | | | | | | | | | | | | | | | | | | | | |
| | Alpine chamois | | | | | | | + | | | | | + | + | | + | | | | | + | + | |
| | Pyrenean chamois | | | + | | | | | + | + | | | + | + | | + | + | | | | + | + | |
| | Wild goat | | + | + | | | | + | | | | | | | | | | | | | + | + | |
| | Bighorn-sheep | | + | | | | | | | | | | | | | | | | | | | + | + |
| | Sheep | | + | | | | | | | | | | | | | | | | | | | + | + |
| Bovinae | Domestic cattle | | + | + | | | + | + | | | | | | + | + | + | + | | | | | | |
| | African buffalo | | + | | | | | | | | + | | | | | + | | + | | | | | |
| | Water buffalo | | + | | | | | | | | | | | | | + | | + | | | | | |
| Alcelaphinae | South African antelope | | + | | | | + | + | + | | | | | + | + | + | + | + | + | + | | + | |
| Cervidae | Reindeer | | | | + | + | | + | + | | | | | | | + | | | | | | + | + |
| | White-tailed deer | | + | + | + | | | | | | | | | | | | + | | | | | + | + |
| | Moose | | + | | | | | | | | | | | | | | + | | | + | + | | + |

Positions in bold are PBR sites according to the human (HLA) DRB1 PBR sites (Brown *et al.*, 1993).

non-synonymous substitutions per non-synonymous site (d_N) (plus s.e.'s) estimated for the exon 2 are given in Table 1.

Analysis of population selection, mutation and recombination

Estimates of population selection (S), mutation (θ) and recombination (ρ) show that each of these processes has

played a significant role in generating the diversity seen among ungulate DRB alleles (Table 3). The likelihood-permutation test, implemented in the LDhat program, indicated the estimates of ρ were significantly different from zero for all taxa in the study (Table 3). With the exception of Spanish ibex, which had a small sample size, all the results were returned with P -values <0.01. Estimates of recombination rate are large for almost all species, and in three cases, domestic sheep, white-tailed

Table 3 Statistics and probability (*P*) values for population selection ($S = 2Ns$; Takahata *et al.*, 1992), population mutation (Watterson's $\theta = 4N\mu$) and population recombination ($\rho = 4Nr$; McVean *et al.*, 2002)

| Group | Species | No. of alleles | R_M | r^2 | P (r^2) | S | θ | ρ | ρ/θ S/ θ | No. of sites | P (LPT) |
|--------------|------------------------|----------------|-------|--------|-------------|-------|----------|--------|------------------------------|--------------|---------|
| Caprinae | Spanish ibex | 6 | 7 | -0.022 | 0.3 | 735.3 | 20.6 | 49 | 2.4 35.7 | 43 | 0.047 |
| | Himalayan tahr | 5 | 6 | -0.211 | <0.0001 | 303.5 | 15.8 | 39 | 2.5 19.2 | 33 | <0.0001 |
| | Bighorn-sheep | 16 | 7 | -0.052 | 0.2 | 391.2 | 11.4 | 13 | 1.1 34.3 | 39 | 0.01 |
| | Alpine chamois | 19 | 7 | -0.211 | <0.0001 | 537.3 | 3.7 | 37 | 9.9 145.2 | 13 | <0.001 |
| | Pyrenean chamois | 13 | 9 | -0.119 | 0.03 | 388.4 | 8.57 | 78 | 9.1 45.3 | 27 | <0.001 |
| | Wild goat | 22 | 17 | -0.080 | 0.04 | 232.7 | 8.6 | 86 | 10 27.1 | 31 | <0.0001 |
| | Sheep | 29 | 11 | -0.131 | 0.01 | 354.1 | 15.3 | >100 | >19.2 23.1 | 21 | <0.0001 |
| Bovinae | Domestic cattle | 20 | 9 | -0.141 | <0.0001 | 310.0 | 8.2 | 66 | 8.0 37.8 | 29 | <0.0001 |
| | Water buffalo | 8 | 10 | -0.123 | 0.001 | 276.7 | 20.8 | 23 | 1.1 13.3 | 54 | <0.0001 |
| | African buffalo | 8 | 10 | -0.164 | <0.0001 | 199.8 | 15.4 | 79 | 5.1 12.9 | 40 | <0.0001 |
| Alcelaphinae | South African antelope | 27 | 6 | -0.095 | 0.03 | 570.0 | 6.5 | 7 | 1.1 87.7 | 25 | 0.009 |
| Cervidae | White-tailed deer | 18 | 13 | -0.093 | 0.014 | 473.0 | 9.0 | >100 | >11.1 52.5 | 31 | <0.0001 |
| | Reindeer | 9 | 6 | -0.316 | <0.0001 | 237.0 | 11.4 | 27 | 2.4 20.8 | 31 | <0.0001 |
| | Moose | 11 | 2 | -0.425 | <0.0001 | 12.6 | 3.4 | >100 | >29.4 3.7 | 10 | <0.0001 |

LPT = *P*-values obtained by likelihood-permutation test implemented in the LDhat program; no. of alleles = number of observed distinct *DRB* sequence alleles for each species; no. of sites = number of segregating biallelic SNPs; R_M = minimum number of inferred recombination events. Alpine ibex is excluded.

deer and moose exceed the maximum value for ρ assessed ($\rho > 100$). The two alternative statistical tests for recombination (r^2 , Awadalla *et al.* (1999); G4, Meunier and Eyre-Walker (2001)) also detected recombination rates significantly different from zero in most cases. For r^2 this was in all species, except for Spanish ibex and bighorn sheep (Table 3, data for G4 are not shown). The DnaSP recombination test identified recombination events (R_M) in all species (Table 3). The ratios ρ/θ and S/θ are also given in Table 3 as an indicator of the likelihood of nucleotides being involved in recombination and selection, respectively, relative to point mutation, and these should be less sensitive to the influence of population size. They show a similar pattern to that exhibited by ρ and S .

Discussion

In this study, we examined genetic diversity in the MHC class II *DRB* gene and demonstrated, via several complementary analysis approaches, that recombination has made a significant contribution to allelic diversity in a range of ungulate species, in addition to the expected role of positive selection. While a role for recombination in generating MHC diversity has previously been suspected, very few studies have been able to provide quantitative estimates of the extent of recombination

(e.g., Richman *et al.*, 2003a,b). In combination with the previous results for rodents, this study suggests that recombination is likely to play a key role in generating MHC diversity for a wide range of taxa, with important implications for interpretation of parameters relating to molecular evolution or inferences that rely on genealogies of MHC alleles.

In common with most other taxa with well-characterized MHC loci, we identified the signature of strong positive selection acting on *DRB* alleles, with the ω ratio > 1 (d_N/d_S ; Yang and Bielawski, 2000) in all the ungulate species. Bayesian analysis identified the individual positively selected sites as being congruent with human HLA-*DRB1* PBR sites (Brown *et al.*, 1993), and the sites were largely consistent across different species. The important role of selection is also indicated by the very high values of S and S/θ (Table 3), obtained for all species except moose. It is tempting to try and assess the relative importance of selection, recombination and mutation in generating *DRB* diversity, by comparing values of S , ρ and θ . However, this is problematic for a number of reasons discussed below. Despite this, the relative ranking of $S > \rho > \theta$ is consistent for all species except moose (which is known to have undergone a severe recent bottleneck), tentatively suggesting the history of most species has been shaped by a similar pattern of evolutionary processes.

Effects of demographic history

The interpretation of these results must be done in the context of the demographic history of each species since changes in population size potentially will have an important influence on the magnitude of parameter estimates and how they interact. Ungulates are a large and diverse group of mammals with over 200 species. The natural distribution covers all continents, with the exception of Australia and Antarctica. Most ungulates have a polygynous mating system, and females usually raise offspring alone. The species included in this study have experienced a range of different demographic histories, for example, buffaloes are likely to have had relatively stable, large population sizes, while alpine ibex have recently experienced an extreme bottleneck. The pattern of sequence variation we observe is consistent with such differing demographic histories, as revealed by the remarkable differences in nucleotide diversity, and in particular silent variation, among the 15 studied ungulates (Table 1), ranging from $d_s = 0.006$ for Alpine chamois to $d_s = 0.106$ for water buffalo. Moreover, for many species low levels of silent variation are present despite high levels of non-silent variation, for example, in Pyrenean ibex and white-tailed deer. Such patterns can arise owing to the interaction between balancing selection and recombination during and after population bottlenecks, as in combination these processes act to distribute non-silent variation among alleles more effectively than silent variation Schierup *et al.* (2001). While both silent and non-silent variation is exposed to the bottleneck, silent variation is less likely to be sampled and make it through the population contraction (Richman *et al.*, 2003a), as selection will favor alleles with functional differences, not silent variation. Thus, the pattern of loss of variation in the DRB gene by Alpine ibex is consistent with the severe bottleneck this species experienced due to over-hunting throughout the 16th and 18th centuries, causing a precipitous decline that brought the species close to extinction (Maudet *et al.*, 2002). Comparable low amounts of MHC class II DRB variation were found in other ruminant species, which experienced severe bottlenecks. For example, in Arabian oryx (*Oryx leucoryx*), only three different exon 2 alleles have been identified at the homologous locus (Hedrick *et al.*, 2000). In Roe deer (*Capreolus capreolus*), in samplings from Norway and Sweden, only four alleles were found, and in moose from northern Europe and North America, 10 different DRB exon 2 alleles have been identified with probably no silent diversity (Mikko and Andersson, 1995; Mikko *et al.*, 1999). The latter study also revealed that musk-ox (*Ovibos moschatus*) and fallow deer (*Dama dama*) were virtually monomorphic at the DRB locus. However, such low levels of diversity are not representative for ungulates, as for other vertebrates, high levels of nucleotide diversity in the MHC are the norm (Mikko *et al.*, 1999).

Limitations of current statistical methodologies

We have taken a comparative approach and tried to estimate the effective population size-scaled parameters (S , ρ , θ) that quantify sequence diversity due to selection, recombination and mutation. A major issue with these methods is that while they are currently among the best-explored approaches available in the literature, they rely

on important assumptions such that ρ is only an estimator of the absolute population recombination rate in the absence of selection, and S is only an estimator of absolute population selection in the absence of recombination. Therefore, obtaining absolute estimates for each parameter under the action of the other is problematic. This is compounded by the fact that we lack methods for estimating confidence intervals for these parameters. While this is highly desirable, for ρ and θ at least, it is an intractable problem under the current theoretical framework used to estimate them (McVean *et al.*, 2002). Further, there is the additional complication that there are likely to have been fluctuations in effective population size for most of the species. Therefore, we are left in a position where it is not easy to be confident of the precision of the absolute values of our parameter estimates. However, results from a range of simulation studies (Satta, 1997; Fearnhead and Donnelly, 2001; Richman *et al.*, 2003b; Smith and Fearnhead, 2005) indicate relative values may be more robust to model violations, and it is these in which we are most interested. We, therefore, argue that we can be confident that there is good evidence for selection and recombination shaping sequence diversity in ungulate DRB genes, as can be best determined with these methods. Similar caution should be applied to studies of other taxa likely to be subject to such biases with these methods.

The impact of violation of model assumption on S is perhaps least well explored (Satta, 1997; Schierup *et al.*, 2001), and differences in potential biases are likely to arise under different demographic histories. Firstly, Takahata *et al.*'s model (1992) has a number of restrictive assumptions, which are not necessarily met by real data, in particular, the assumption of a fixed number of alleles through time. The model also assumes an infinite allele model where mutation occurs through the substitution of single nucleotides, whereas recombination may generate alleles that differ at multiple positions relative to the parental alleles. Biases in S can then arise through the effect recombination may have on the number of alleles segregating in a population and the number of nucleotide differences between them. Under the model assumptions of Takahata *et al.*, the number of alleles in the population should be approximately equal to K_n , the mean number of pairwise non-synonymous differences in the PBR among sequences (hence S can be approximated using K_n calculated from the data). However, the simulation study of Satta (1997) showed that recombination can greatly increase the number of alleles in a population (n), while decreasing K_n (since recombination will ultimately act to homogenize variation along the length of the sequence). In consequence, recombination may act to bias S upwards if calculated directly from n , but could bias it downwards slightly if n is estimated from K_n . This effect will show a strong interaction with demographic history, as bottlenecks in particular can have a substantial effect on the number of alleles and sequence diversity among them. If bottlenecks, in conjunction with recombination, remove synonymous variation at loci under balancing selection (Schierup *et al.*, 2001; Richman *et al.*, 2003a), then in these circumstances S could be overestimated. Such a pattern may be visible in our data where the highest values of S were obtained for Spanish ibex and Alpine chamois, species known to have experienced recent bottlenecks.

Additional sources of error come from potential misidentification of PBR sites in these taxa, functional significance of non-synonymous variation outside of the PBR and lack of correction for multiple hits. These could all potentially bias S downwards as they would lead to an underestimation of K_n . Recombination is also likely to be an important consideration in the estimation of the magnitude of positive selection acting on a sequence and the identification of individual sites under selection using maximum-likelihood phylogenetic approaches (Anisimova *et al.*, 2003; Wilson and McVean, 2006), as with recombination there will be multiple evolutionary trees along the sequence. This can lead to an upward bias of ω and false positives for sites under selection. Having said this, the methods used in the current study do appear to be relatively robust (Anisimova *et al.*, 2003). Ultimately, while caution must be used in assessing the precision of our estimators of selection, the fact that strong signatures of selection are obtained from different methods lets us be reasonably confident that selection has been an important process in shaping variation for these sequences.

In a recent simulation study, Smith and Fearnhead (2005) evaluated the accuracy and robustness of three population-scaled recombination rate estimators, including that calculated by LDhat. They concluded that while absolute values were likely to be sensitive to violations of standard model assumptions, relative recombination rates were inferred reasonably well. Although a balancing selection scenario was not included in their study, they did examine bias due to a selective sweep, finding only a weak influence on all three methods, with the approach used by LDhat showing $\sim 2\%$ upward bias. They also found there was some evidence that bottlenecks may bias estimates of recombination downwards by $\sim 35\text{--}45\%$, and population growth can lead ρ to be overestimated by $10\text{--}30\%$. These demographic influences seem plausible since in very low-diversity populations recombination is less likely to generate new allelic variants than in cases where some variation is retained. Further support to the robustness of this inference approach comes from the fact that ρ values obtained by LDhat have been shown to be in strong concordance to ρ estimates derived from pedigree studies and from sperm-typing experiments (Myers *et al.*, 2005). Thus, our observations of significant recombination rates are likely to be robust, although again the absolute values should be interpreted with caution. Similarly, Fearnhead and Donnelly (2001) demonstrated that the ratio ρ/θ is also robust to violations of the standard model, and should be less sensitive to variation in N , lending some support to our observation that overall, recombination rates are likely to exceed mutation rates for ungulates (although the absolute values should be treated cautiously) and, therefore, that potentially recombination may play a more significant role than mutation in generating allelic diversity at MHC loci for these species.

Clearly, there is need for progress in this area and recently Wilson and McVean (2006) have reported a new method for the simultaneous estimation of ω and ρ using a coalescent model implemented in a Bayesian framework. Their initial application of this method to sequences for *porB3*, a protein thought to be important for pathogenesis in *Neisseria meningitidis*, appears to have

strong potential for overcoming many of the issues raised here, but its general utility remains to be explored. Integrating the quantitative interpretation of these statistics with fluctuations in demography remains problematic however.

General conclusions

Recombination is a powerful molecular mechanism that may generate (under balancing selection) genetic diversity in the MHC genes. The functional constraints within the exon 2 are restricted to relatively few positions (e.g., cysteine residues at positions 15 and 79 for correct folding), so high mutation and recombination rates may only rarely have negative consequences for the molecule's function. In populations that have undergone bottlenecks, but which have retained some diversity, allelic variation may be restored much faster with recombination than under the occurrence of point mutations alone. Further comparative studies of recombination rates at MHC loci in diverse taxa will aid in assessing the general importance of recombination in generating MHC variation. These are likely to add to the growing number of theoretical and empirical studies indicating that recombination plays a key role in generating MHC diversity, which is then maintained by selection (Satta, 1997). Future improvements to theoretical frameworks for selection recombination rate estimators (e.g., Wilson and McVean, 2006) will allow more robust assessments of interactions between selection, recombination and demographic history, and how together they shape patterns of allelic diversity at the MHC, or other loci subject to strong balancing selection.

One important consequence of recombination in the *DRB* gene is the effect on the shape of inferred genealogies. Schierup and Hein (2000) showed in their study that even very low rates of recombination significantly alter the shape of the inferred phylogenetic trees. They concluded that ignoring recombination in phylogenetic-based analyses of sequence data from populations causes underestimation of the time both to the most recent common ancestor and the amount of recent divergence. Moreover, it leads to the loss of a molecular clock and generates more apparent ancient polymorphism (e.g., trans-specific evolution). Thus, the origin of ungulate MHC class *DRB* alleles may be, in some cases, much more recent than previously assumed. Consequently, if similar levels of recombination occur in the MHC genes of other taxa as observed in this study, the use of traditional phylogenetic methods and inferences made from them may be inadequate.

Acknowledgements

We thank Marco Giacometti from the Wild vet Projects, Stampa for organizing and providing Alpine ibex samples and Günther Hartl from the Christian-Albrechts-Universität zu Kiel for the Spanish ibex samples. We also thank Thorsten Reusch from the Max-Planck Institute of Limnology for his critical reading of an early draft of the manuscript. Finally, we would like to thank the editor and the anonymous reviewers whose comments improved the quality of the final paper.

References

- Amills M, Jimenez N, Jordana J, Riccardi A, Fernandez-Arias A, Guiral 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.
- Andersson L, Mikko S (1995). Generation of MHC class II diversity by intra- and intergenic recombination. *Immunol Rev* **143**: 5–12.
- Anisimova M, Nielsen R, Yang Z (2003). Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**: 1229–1236.
- Awadalla P, Eyre-Walker AA, Maynard Smith J (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**: 2524–2525.
- Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998). Recent origin of HLA-DRB1 alleles and implications for human evolution. *Nat Genet* **18**: 237–242.
- Bernatchez L, Landry C (2003). MHC studies in non-model vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* **16**: 363–377.
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL *et al.* (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **64**: 33–39.
- Doherty PC, Zinkernagel RM (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**: 50–52.
- Fearnhead P, Donnelly P (2001). Estimating recombination rates from population genetic data. *Genetics* **159**: 1299–1318.
- Figueroa F, Gunther E, Klein J (1988). Mhc polymorphism predating speciation. *Nature* **335**: 265–267.
- Gyllensten UB, Sundvall M, Erlich HA (1991). Allelic diversity is generated by intraexon sequence exchange at the DRB1 locus of primates. *Proc Natl Acad Sci USA* **88**: 3686–3690.
- Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 96/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Hedrick PW, Parker KM, Gutierrez Espeleta GA, Rattink A, Lievers K (2000). Major histocompatibility complex variation in the Arabian oryx. *Evolution* **54**: 2145–2151.
- Hudson RR (2001). Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- Hudson RR, Kaplan NI (1985). Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* **111**: 336–347.
- Hughes AL, Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature* **335**: 167–170.
- Hughes AL, Yeager M (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* **32**: 415–435.
- Hughes AL (2002). Natural selection and the diversification of vertebrate immune effectors. *Immunol Rev* **190**: 161–168.
- Jordan WC, Bruford WM (1998). New perspectives on mate choice and the MHC. *Heredity* **81**: 127–133.
- Jukes TH, Cantor CR (1969). Evolution of protein molecules. In: Munro HN (ed). *Mammalian Protein Metabolism*. Academic Press: New York, pp 21–132.
- Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.
- Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER *et al.* (1990). Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31**: 217–219.
- Klein J, Takahata N, Ayala FJ (1993). Mhc diversity and human origins. *Sci Am* **269**: 46–51.
- Kumar S, Tamura K, Jakobsen I, Nei M (2001). *MEGA 2.1: Molecular Evolutionary Genetics Analysis Software. Version 2.1*. Arizona State University: Tempe, AZ, USA.
- Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T (2002). Experimental evidence for major histocompatibility complex–allele-specific resistance to a bacterial infection. *Proc Roy Soc Lond Ser B* **269**: 2029–2033.
- Maudet C, Miller C, Bassano B, Breitenmoser-Würsten C, Gauthier D, Obexer-Ruff G *et al.* (2002). Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex (*Capra ibex (ibex)*). *Mol Ecol* **11**: 421–436.
- McVean GAT, Awadalla P, Fearnhead P (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004). The fine-scale structure of recombination rate variation in the human genome. *Science* **304**: 581–584.
- Meunier J, Eyre-Walker A (2001). The correlation between linkage disequilibrium and distance. Implications for recombination in Hominid mitochondria. *Mol Biol Evol* **18**: 2132–2135.
- Mikko S, Andersson L (1995). Low major histocompatibility complex class II diversity in European and North American moose. *Proc Natl Acad Sci USA* **92**: 4259–4263.
- Mikko S, Roed K, Schmutz S, Andersson L (1999). Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. *Immunol Rev* **167**: 169–178.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**: 321–324.
- Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.
- Nielsen R, Yang Z (1998). Likelihood models for detecting positively selected amino-acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.
- Paterson S, Wilson K, Pemberton JM (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proc Natl Acad Sci USA* **95**: 3714–3719.
- Penn DJ, Potts WK (1999). The evolution of mating preferences and major histocompatibility complex genes. *Am Nat* **153**: 145–164.
- Penn DJ, Damjanovich K, Potts WK (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci USA* **99**: 11260–11264.
- Ptak SE, Roeder AD, Stephens M, Gilad Y, Pääbo S, Przeworski M (2004). Absence of the TAP2 human recombination hotspot in chimpanzees. *PLOS Biol* **6**: 0849–0855.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Reusch TBH, Häberli MA, Aeschlimann PB, Milinski M (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* **414**: 300–302.
- Richman AD, Herrera LG, Nash D (2003a). Evolution of MHC class II E β diversity within the genus *Peromyscus*. *Genetics* **164**: 197–289.
- Richman AD, Herrera LG, Nash D, Schierup HM (2003b). Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. *Genet Res Camb* **82**: 89–99.
- Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ *et al.* (2003). IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* **31**: 311–314.
- Satta Y (1997). Effects of intra-locus recombination on HLA polymorphism. *Hereditas* **127**: 105–112.
- Schaschl H, Goodman SJ, Suchentrunk F (2004). Sequence analysis of the MHC class II DRB alleles in Alpine chamois (*Rupicapra r. rupicapra*). *Dev Comp Immunol* **28**: 265–277.

- Schaschl H, Suchentrunk F, Sabine Hammer, 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.
- Schierup MH, Hein J (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.
- Schierup M, Mikkelsen AM, Hein J (2001). Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics* **159**: 1833–1844.
- She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK (1991). Amplification of major histocompatibility complex class-II gene diversity by intraexonic recombination. *Proc Natl Acad Sci USA* **88**: 453–457.
- Smith NGC, Fearnhead P (2005). A comparison of three estimators of the population-scaled recombination rate: accuracy and robustness. *Genetics* **171**: 2051–2062.
- Stumpf MPH, McVean GAT (2003). Estimating recombination rates from population-genetic data. *Nat Rev Genet* **4**: 959–968.
- Takahata N, Nei M (1990). Allelic genealogy under over-dominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967–978.
- Takahata N, Satta Y (1998). Selection, convergence, and intragenic recombination in HLA diversity. *Genetica* **157**: 157–169.
- Takahata N, Satta Y, Klein J (1992). Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**: 925–938.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**: 4876–4882.
- Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003). Parasites selection for immunogenetic optimality. *Science* **301**: 1343.
- Wilson DJ, McVean G (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**: 1411–1425.
- Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Cabios* **13**: 555–556.
- Yang Z, Bielawski (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**: 496–503.
- Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.