Immunogenetics (2008) 60:585–598
DOI 10.1007/s00251-008-0319-x

ORIGINAL PAPER

Polymorphism, haplotype composition, and selection in the *Mhc-DRB* of wild baboons

Elise Huchard · Mylene Weill · Guy Cowlishaw · Michel Raymond · Leslie A. Knapp

Received: 22 April 2008 / Accepted: 30 June 2008 / Published online: 5 August 2008

Abstract General patterns of organization in the major histocompatibility complex (MHC) have been successfully explained by the model of birth-and-death evolution, but understanding why certain MHC genes are maintained together into specific haplotypes remains challenging. The haplotype configurations of the functionally important class II DR region have been described in few primates and display important interspecific variability with respect to the extent of allelic variation, the number of loci and/or combinations of loci present. Understanding the evolutionary mechanisms driving such variation is conditional upon characterizing haplotypes in new species and identifying the selective pressures acting on haplotypes. This study explores the variability of haplotype configurations in the *Mhc-DRB* region (exon 2) for the first time in wild non-human primates, chacma baboons (*Papio ursinus*). *Paur-DRB* haplotypes were characterized through segregation studies and linkage disequilibrium. 23 *Paur-DRB* sequences and 15 haplotype configurations were identified in 199 animals. The *Paur-DRB* exon 2 is shown to be subjected to intense positive selection and frequent recombination. An approach recently developed for human vaccine studies was used to classify *Paur-DRB* sequences into supertypes, based on the physico-chemical properties of amino acids that are positively selected, thus most probably involved in antigen recognition. Sequences grouped into the same supertype (thus presumably sharing antigen-binding affinities) are non-randomly distributed within haplotypes, leading to an increased individual diversity of supertypes. Our results suggest that selection favoring haplotypes with complementary sets of *DRB* supertypes shapes functionally tuned haplotypes in this natural baboon population.

Keywords Major histocompatibility complex · Primates · Haplotypes · Polymorphism · *Papio ursinus*

E. Huchard (✉) · M. Weill · M. Raymond
CNRS, Institut des Sciences de l'Evolution,
Université Montpellier 2,
CC 065, Place Eugène Bataillon,
34095 Montpellier Cedex 05, France
e-mail: ehuchard@univ-montp2.fr

E. Huchard · G. Cowlishaw
Institute of Zoology, Zoological Society of London,
Regent's Park,
London NW1 4RY, UK

L. A. Knapp
Department of Biological Anthropology,
University of Cambridge,
Downing Street,
Cambridge CB2 3DZ, UK

Introduction

Gene duplication plays an important role in the adaptive evolution of vertebrates and specific patterns of duplication have been suggested for many different multi-gene families, including the major histocompatibility complex (MHC). The model of birth and death evolution for the MHC gene family was originally proposed by Nei and Hughes (1992), who suggested that new MHC genes are created by repeated duplications that may ultimately be deleted from the genome or could instead accumulate mutations that render them either inactivated pseudogenes or novel alleles with new selective advantages. This model is extremely powerful in its ability to explain general patterns of MHC

organization in most vertebrates (e.g., Gu and Nei 1999; Nei et al. 1997), including primates (Piontkivska and Nei 2003). However, it does not offer insight into why certain MHC genes are maintained together in the form of specific haplotypes. In an attempt to answer this question and to gain a better understanding of the evolutionary strategies used to generate and maintain MHC variation in primates, we believe it is necessary to evaluate MHC diversity at the genomic and haplotype, as well as the allelic level, whenever possible.

The MHC is the most variable part of the vertebrate genome and has been intensively studied in primates over the last two decades. The human leukocyte antigen (HLA) system is the most thoroughly studied primate MHC, located on the short arm of chromosome 6. The HLA class II region is divided into the *HLA-DR*, *HLA-DQ*, and *HLA-DP* regions, which each contain multiple A (alpha) or B (beta) loci. The classical *HLA-DR*, *HLA-DQ*, and *HLA-DP* molecules are transmembrane dimers, composed of an alpha and beta chain subunit encoded by the *A* and B genes, respectively. MHC class II molecules are specialized in the presentation to T-helper cells of peptides processed by specialized antigen presenting cells from endocytosed (phagocytized) antigens. The antigen-specific T-helper cells promote and regulate the development of antigen-specific cellular and humoral responses, leading to the production of effectors that have in charge the destruction of the antigen. Parts of the functionally important antigen-binding sites (ABS) are encoded by the second exon of MHC class II *DRB* molecules (Brown et al. 1993), and genetic diversity can be extremely high at the population level, both in terms of numbers of alleles and extent of sequence diversity among alleles (Apanius et al. 1997). MHC variability, which allows binding of a wide array of antigens, may be maintained by some form of pathogen-driven balancing selection (heterozygote advantage or frequency-dependent selection; reviewed by Hedrick and Kim 2000; Hughes and Yeager 1998; Jeffery and Bangham 2000).

The *HLA-DR* region contains a *DRA* gene exhibiting limited polymorphism and a differential number of *DRB* loci that have been designated *HLA-DRB1* to *HLA-DRB9* (Bodmer 1972). Four of these loci, *DRB1*, *DRB3*, *DRB4*, and *DRB5* are functional and capable of presenting peptides to T cells. The remaining loci, *DRB2*, *DRB6*, *DRB7*, *DRB8*, and *DRB9*, are pseudogenes, containing one or more structural abnormalities that render them non-functional. Functional and orthologous *DRB1*, *DRB3*, *DRB4*, and *DRB5* loci were described in several non-human primate species (reviewed in Slierendregt et al. 1992).

While the organization of the *DRB* region of chimpanzees is highly similar to its human equivalent (Bontrop et al. 1999), the organization of the well-studied *Mamu-DRB* region of rhesus macaques (*Macaca mulatta*) is very different and displays variation at the population level with regard to the number and/or combination of loci present per configuration (Doxiadis et al. 2001; Doxiadis et al. 2000; Otting et al. 2000; Slierendregt et al. 1994). One to eight *DRB* loci can be found in the rhesus macaque haplotype, whereas no more than five loci are present in human haplotypes (Bontrop et al. 1999; Doxiadis et al. 2001). Up to 31 *Mamu-DRB* haplotype configurations have been described so far (Doxiadis et al. 2001), which contrasts with the five *HLA-DRB* region configurations. Moreover, each *HLA-DRB* configuration displays a high degree of polymorphism (Marsh 2005; Robinson et al. 2003). Although the total number of *Mamu-DRB* sequences is comparable to those assigned to the *HLA-DRB1* locus, the *Mamu-DRB* haplotypes show only a limited degree of allelic variation (Doxiadis et al. 2000; Khazand et al. 1999).

The evolutionary significance of the extensive interspecific variability encountered in the primate *Mhc-DRB* region remains poorly understood, as preliminary attempts to address this question mainly relied on circumstancial evidence arising from comparative observations (e.g., Bontrop et al. 1999). This is also due to a lack of empirical data. Although MHC sequences are now available for a wide range of species, data describing the haplotype configurations of the *DR* region are still limited to a handful of primate species (e.g., Bontrop et al. 1999). Progress in this area is thus conditional upon (1) characterizing the *Mhc-DRB* region in closely related primate species and (2) gaining insights into the selective pressures acting on the number and associations of loci within region configurations or haplotypes. While the former step appears essentially descriptive, efforts should be made to take into account the distinct functional (i.e., binding) abilities of the different *Mhc-DRB* alleles to tackle the latter step, since allelic polymorphism is likely to be maintained through balancing selection.

Considerable progress has recently been made in understanding molecular events involved in the processing and presentation pathways utilized by MHC molecules. Crystallographic structures of several peptide-MHC molecules have been determined, and sequence motifs associated have been derived (Madden 1995; Stern et al. 1994). Certain class I alleles have been found to recognize similar antigen motifs and were thus grouped into HLA supertypes (Delguercio et al. 1995; Sette and Sidney 1998; Sette and Sidney 1999; Sidney et al. 1996). The classification of MHC molecules into supertypes, based on common structural and functional features, is useful in the development of epitope-based vaccines (Sette et al. 2001, 2002). Therefore, bioinformatic approaches have been proposed to overcome the prohibitive cost of an experimental determination of

antigen motifs bound by each allele (Doytchinova et al. 2004; Lund et al. 2004) and were applied to class II MHC molecules (Doytchinova and Flower 2005). By grouping alleles with a potential functional overlap that could be subject to identical pressures, a MHC supertype could also be considered as a unit of selection.

This approach represents a new avenue to investigate MHC selection in natural populations, where the extreme polymorphism requires unrealistic sample sizes to identify allele-specific effects. Recent field studies have for instance been able to show that MHC supertypes can successfully predict HIV progression in humans (Trachtenberg et al. 2003) and parasite loads in fat-tailed dwarf lemurs (Schwensow et al. 2007). Whereas MHC supertypes can be considered as selective units, MHC haplotypes are regions of nucleotide sequence (blocks) characterized by strong linkage disequilibrium and constitute a transmission unit from parent to offspring. Some haplotypes consist of more than 3 Mb in humans (Alper et al. 2006), whereas the HLA class II region, divided into the *HLA-DR*, *HLA-DQ*, and *HLA-DP* regions, encompasses around 700,000 base pairs. Consequently, a MHC haplotype is targeted by a number of possibly divergent selective pressures resulting from allele-specific effects. Using supertype classification to approximate specific binding affinities of *Mhc-DRB* alleles may prove useful to explore the selective processes influencing the number and associations of loci present in *Mhc-DRB* haplotypes.

An extensive survey from a natural population of chacma baboons (*Papio ursinus*) provides an opportunity to improve our understanding of the *Mhc-DRB* organization in an African cercopithecine primate, while most of the current knowledge regarding *Mhc-DRB* organization was gathered from macaques (especially *M. mulatta* and *M. fascicularis*). In our previous study of *Mhc-DRB* variation in 18 chacma baboons, we found sequences belonging to *DRB1*, *DRB5*, and 13 other lineages (Huchard et al. 2006). The present study aims at extending this initial survey by describing both sequence polymorphism and *Mhc-DRB* haplotypes in 199 animals. Formally stated, our objectives are straightforward. First, we aim to investigate, by providing a first characterization of *Mhc-DRB* haplotypes in *P. ursinus*, whether baboon *Mhc-DRB* region configuration and sequence variation is more similar to the macaque or human *Mhc-DRB* region. Second, we aim to estimate the intensity of diversifying selection in baboon *Mhc-DRB* sequences and to classify sequences into supertypes. Finally, we aim to test if *Mhc-DRB* sequences are randomly associated within haplotypes with respect to their structure and/or function. This test, in particular, represents an important step toward an understanding of the evolutionary processes underlying divergence in the organization of primate *Mhc-DR* region.

Materials and methods

Study population and tissue sampling

For *Mhc-DRB* genotyping, genomic DNA was extracted from tissue samples from 199 wild chacma baboons (*P. ursinus*) living in Tsaobis Leopard Park, on the edge of the Namib Desert in Namibia, Southern Africa (for details of the site and population, see e.g. Cowlishaw 1999). Beginning in 2000, the baboons were captured in order to gather biological samples and morphological measurements. Tissue samples were obtained from 199 individuals, including two solitary males and individuals belonging to six troops during trapping operations in 2000, 2001, 2005, and 2006. Our trapping procedure captured an average of 100% of adult males and 90% of adult females from a troop (estimations based on eight trapping operations, unpublished data). Among these individuals, 56 mother-offspring pairs were identified using behavioral observations and confirmed with microsatellite data (Knapp et al., unpublished). Tissue samples were stored in a dimethyl sulphoxide–salt solution (Seutin et al. 1990), and DNA was extracted using a DNeasy Tissue Kit (Qiagen, Crawley, UK) following the manufacturer's instructions.

Mhc-DRB genotyping

Paur-DRB sequences (for *P. ursinus*) were obtained by polymerase chain reaction (PCR) from genomic DNA, using generic primers that amplify multiple sequences spanning most of exon 2 and including the entirety of the peptide-binding region. We have previously shown that our methods amplify multiple loci in *P. ursinus* and have assessed sequence heterogeneity of amplification products using denaturing gradient gel electrophoresis (DGGE; Huchard et al. 2006). Details on the molecular methods involving DGGE are also available in Myers et al. (1987) and Knapp et al. (1997). Bands on DGGE gels were excised, reamplified using original primers, and directly sequenced in both directions using an ABI prism 310 sequencer and the Big Dye Terminator kit (Invitrogen). We controlled PCR artifacts by sequencing each new band at least three times from different individuals. When a sequence was recovered from a single individual in the sample, sequencing was performed on multiple (and therefore independent) amplification products. When bands contained several sequences migrating in a same position on a DGGE gel, PCR products were cloned using a Topo TA Cloning kit (Invitrogen) following the manufacturer's instructions. For each individual, at least 20 bacterial colonies were directly amplified with the original primers. Amplification products were then digested independently

with three restriction enzymes known to cut frequently (*Mbo*II, *Hae*III, and *Alu*I) and electrophoresed on 2% agarose gel to establish its restriction profile (see Huchard et al. 2006 for details on the restriction fragment length polymorphism protocols). For each restriction profile, a minimum of three clones were purified with a Qiagen plasmid preparation kit and sequenced as described above using the primers M13R-pUC and M13F (Invitrogen). New sequences obtained from clones were subsequently electrophoresed on a DGGE gel to ensure that they were migrating in the expected position. Fifty-five percent of the samples were electrophoresed more than once on a DGGE gel using independent PCR products from a same individual to ensure the repeatability of our results. Inconsistencies in results from the same individual were resolved by amplifying again and electrophoresing the sample on a new DGGE until consistent results were obtained. No sample was electrophoresed more than three times. New *Paur-DRB* sequences were named by the IMGT-NHP (Immunogenetics Non-human Primate) Nomenclature Committee (Robinson et al. 2003). *Paur-DRB* haplotypes were mainly deduced from segregation analysis within mother–infant pairs (Table 1). These were subsequently confirmed by patterns of linkage disequilibrium (Fig. 1) since *DRB* sequences were non-randomly associated within individuals, with a limited number of combinations (i.e., haplotypes) identified in the whole sample. A haplotype was accepted when identified in at least two individuals.

Tests for positive selection

The presence of positively selected sites (PSS), characterized by $\omega = d_N/d_S > 1$, where d_N and d_S are the relative amounts of substitutions at non-silent and silent codon sites, was investigated in *Paur-DRB* amino acid sequences. We compared the null model (without positive selection), where $\omega<1$ (model M7), and a model allowing an additional class of sites where $\omega>1$ to account for the possible occurrence of PSS (model M8) using a likelihood ratio test (LRT; Yang et al. 2000). If M8 fits the dataset better than M7, PSS are subsequently identified using an empirical Bayes method. Analyses were carried out using the software CODEML, implemented in the package PAML version 3.14 (Yang 1997). M7 and M8 have proved to be more robust toward intragenic recombination than other implemented models, as well as the Bayes' prediction of sites under positive selection (Anisimova et al. 2003).

d_S, d_N, and their standard errors were then estimated using a second approach: the evolutionary pathways method (Nei and Gojobori 1986) implemented in MEGA 3.1 (Kumar et al. 2001), applying the correction of Jukes and Cantor (1969) for multiple hits. In counting the pairwise number of silent and non-silent substitutions in a set of sequences, the evolutionary pathways method considers all possible evolutionary pathways (excluding termination codons) leading from one codon to another as equally probable and therefore makes fewer assumptions than other methods (Nei and Kumar 2000). Separate tests were conducted for sites predicted to be involved in antigen recognition, assuming concordance to human ABS (Brown et al. 1993; Slierendregt et al. 1992) as well as for non-ABS sites, and then for PSS and non-PSS. A global test was also conducted using all sequences to calculate the overall values of d_N and d_S.

Definition of supertypes

MHC supertypes were subsequently defined by applying amino acid sequence-based clustering as proposed by Doytchinova and Flower (2005). In a first step, amino acids from PSS were selected and aligned. Each of them was then described by a vector of five physicochemical descriptor variables: z1 (hydrophobicity), z2 (steric bulk), z3 (polarity), z4, and z5 (electronic effects) (Sandberg et al. 1998), in order to obtain a matrix. This matrix was subjected to hierarchical clustering using average linkage clustering and three similarity measurements (Euclidian distance, Pearson correlation coefficient, and cosine correlation coefficient), using Genesis 1.6.0 Beta 1 (Sturn et al. 2002). Clustering trees obtained using different similarity measurements were slightly different. In the absence of established criterion to define supertypes from a clustering tree using a limited number of sequences (when antigen-binding affinities cannot be assessed through immunoassays), those sequences that were consistently grouped together in all three trees were considered part of the same supertype. We thus obtained similar (consensus) supertypes from different trees. This approach (rather than tracing an arbitrary line in one of the trees) was considered conservative because it minimized the risk of grouping sequences with distinct binding affinities within a same supertype by taking into account only robust nodes of the clustering trees to create supertypes.

Estimation of population recombination rate

The population recombination rate ($\rho=4N_e r$, with r=the per gene per generation rate of crossing over and N_e= the effective population size) characterizing the *Paur-DRB* exon 2 was calculated using Hudson's composite-likelihood estimate approach (Hudson 2001). This method has been extended by McVean et al. (2002) to estimate the population recombination rate conditioned on the estimate of mutation rate per site ($\theta=4N_e\mu$, with μ the mutation rate) to take into account high rates of recurrent mutations in sequences. A likelihood permutation test was used to test

Table 1 *Paur-DRB* sequences and haplotypes identified from 25 baboons, including 17 mother-offspring pairs

Ind.	Mo.	1*0301	1*0302	1*0303	5*0301	5*0302	W101	W301	W401	W2701	W2801	W3601	W4801	W5301	W5501	W5601	W5602	W5701	W5801	W5901	6*0101	6*0103	6*0102	6*0104	*Mhc-DRB* genotype
HF16			E					E	E												E				EE
LF14	HF16		E					E	E	A			A		A						E			A	AE
LM12	HF16		E		B			E	E							B					E				BE
GF14		D			H												H			D					DH
GF15	GF14	C	C		H												H								CH
JF04					B											B									BB
JF03	JF04				B/H											B	H								BH
JF30	JF04		E		E			E	E							B					E				BE
HF22		D			B											B				D					BD
LF10	HF22	D		G			G											G	G	D					DG
LM09	HF22	D								A			A		A					D				A	AD
LF17	HF22				B											B									BB
JF23			E					E	E			F						F			E		F		EF
JM11	JF23			G			G					F						F/G	G				F		FG
JF15	JF23	D										F						F		D			F		DF
JM22	JF23		E		B			E	E							B					E				BE
JM29	JF23				H							F					H	F					F		FH
JF17			E		B			E	E							B					E				BE
JF08	JF17	I	E			I		E	E												E				EI
JF29	JF17		E					E	E			F						F			E		F		EF
LF01					H									K			H								HK
LF03	LF01				H												H								HH
LM15	LF01				H					A			A		A		H							A	AH
HF14					B					A			A		A	B								A	AB
HF13	HF14									A			A	K	A									A	AK

Letters designate co-segregating sequences, i.e., haplotypes, made up of one to four *Mhc-DRB* sequences. .Each sequence is named according to the IMGT-NHP nomenclature

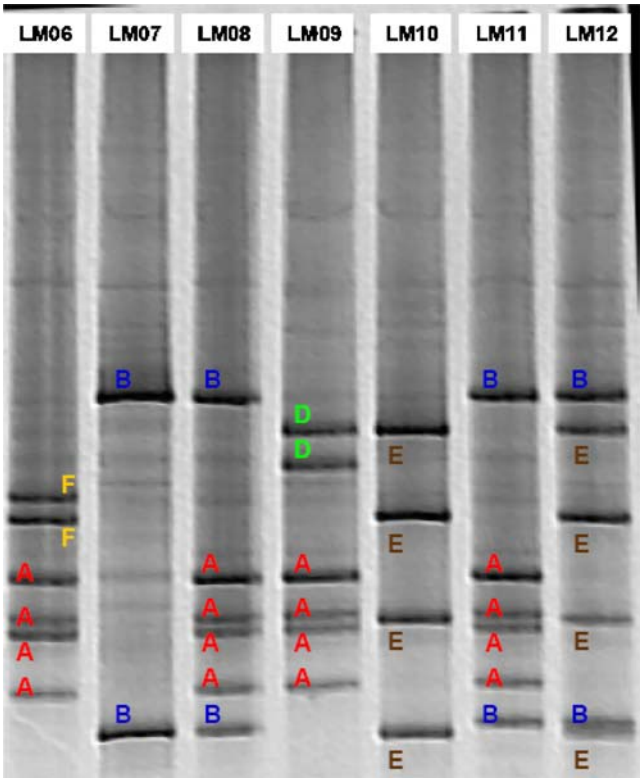

Fig. 1 *Mhc-DRB* haplotype typing using DGGE. Electrophoresis of the PCR product for seven individuals. Each band on the gel corresponds to a specific sequence. *Letters* designate co-segregating bands, i.e., haplotypes. Individuals LM07 and LM10 are homozygous at the *Mhc-DRB* region, thus exhibiting only one haplotype each. See text for explanations

the null hypothesis of no recombination (ρ=0) using the software LDhat (McVean et al. 2002). This should provide accurate estimates of recombination rates with regard to sequences evolving under symmetric balancing selection (Richman et al. 2003b).

Distribution of supertypes within haplotypes

The distribution of *Mhc-DRB* sequences showed no occurrence of functionally overlapping sequences (i.e., grouped into a same supertype) within a same haplotype. In order to investigate whether *Paur-DRB* sequences were non-randomly associated with respect to their function within haplotypes, the probability of this event under the hypothesis of random distribution of supertypes within haplotypes was estimated by randomization. *Mhc-DRB* sequences were randomly allocated to haplotypes. The number of loci per haplotype and individual genotypes (and thus, haplotype frequency in the population) were extracted from the data. For each random sample, the number (Nh) of similar supertypes within a same haplotype was recorded, as well as the number (Ni) of distinct supertypes for each individual. A first *P* value was computed as the ratio of all cases displaying a lower or equal Nh than the observed sample, over the number of random samples (10.000). To test whether the actual distribution of supertypes per haplotype increases the number of supertypes per individual, a second *P* value was computed as the ratio of all cases displaying a larger or equal Ni than the observed sample, over the number of random samples (10.000). In order to control for the potential confusing factor introduced by sequence similarity, the 23 *Paur-DRB* sequences (full length) were also grouped on the basis of their nucleotide and predicted amino acid similarity; thus, two alleles from a same locus/lineage are expected to cluster together. Two neighbor-joining trees were built using MEGA (Kumar et al. 2001), respectively based on (1) the number of nucleotide differences using the nucleotide sequences (252 bp) and (2) the number of amino acid differences using the predicted amino acid sequences (84 aa). In each case, the position of the cutoff line defining clusters aimed to obtain 12 groups to keep a clustering level comparable with the supertype randomization test (Fig. 2). A randomization test was performed as described above for each of these two extra clustering criteria. Sequences belonging to the *DRB6* loci are pseudogenes in humans, chimpanzees, and macaques (Bontrop et al. 1999), thus may be pseudogenes in baboons. The four *Paur-DRB6* sequences were therefore excluded from each set of simulations (for sequences grouped with respect to physico-chemical properties of the PSS, to nucleotide, or to amino acid similarity).

Results

Paur-DRB sequence polymorphism

Nucleotide sequence analysis of 252 bp from exon 2 of the MHC class II *DRB* region of 199 wild chacma baboons revealed 23 distinct *Paur-DRB* sequences. Sixteen sequences were previously reported from this population (Huchard et al. 2006), and seven are new (Genbank/EMBL/DDBJ accession numbers EU244816–EU244822). As mentioned previously, the 252-bp region contains the entire peptide-binding region, and the flanking regions of this exon are highly conserved within and between species. According to IMGT nomenclature, *Paur-DRB* sequences can be assigned to three different primate locus groupings: *DRB1*, *DRB5*, *DRB6*, and to 13 *DRB**W* lineages. Lineages are defined as a cluster of highly related sequences probably originating from a shared ancestral structure (Bontrop et al. 1999). Some lineages identified here have been described for other Old World primates, and some have not been previously described in primates. No stop codons, insertions, or deletions were found. Considering all 23 sequences, a total of 111 variable nucleotide positions were identified, repre-

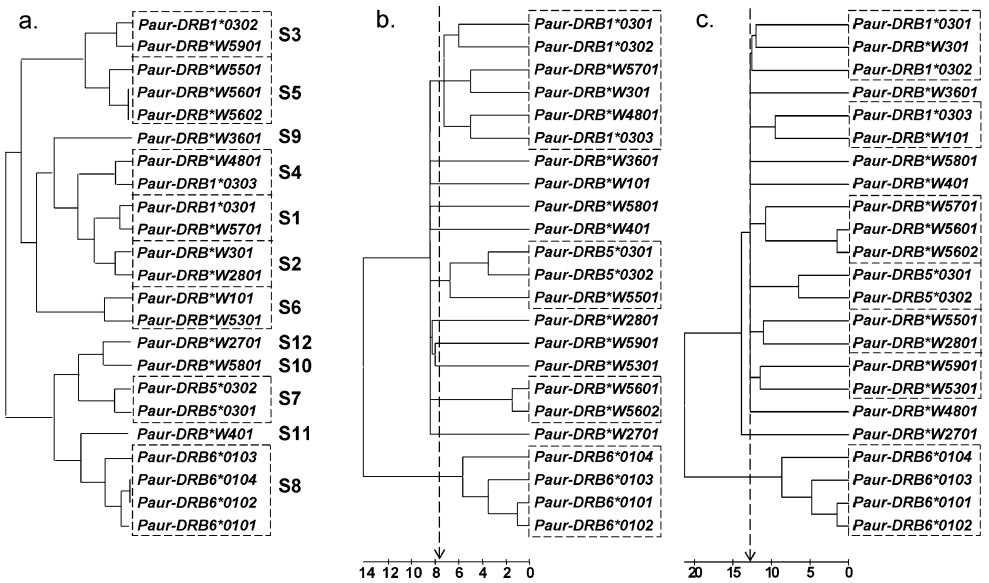

Fig. 2 Comparison of 23 *Paur-DRB* sequences using three different clustering criteria. **a** Tree constructed by hierarchical clustering. *Paur-DRB* sequences were grouped into 12 supertypes (S1–12) derived from amino acid sequences and based on physicochemical properties of positively selected codon sites. **b** Neighbor-joining tree derived from the number of amino acid differences separating *Paur-DRB* predicted amino acid sequences (84 aa). The *dashed arrow* indicates the cutoff line used to define 12 groups of *Paur-DRB* amino acid sequences. **c** Neighbor-joining tree derived from the number of nucleotide differences separating *Paur-DRB* nucleotide sequences (252 bp). The *dashed arrow* indicates the cutoff line used to define 12 groups of *Paur-DRB* nucleotide sequences. See text for details

senting 44% (111/252) of the sequence analyzed. Pairwise comparisons between sequences revealed an average of 32.9 (or 13%) substitutions (range, 3–52 or 1–21%). Each sequence had a distinct predicted sequence of 84 amino acids, exhibiting 51 (60%) variable sites (Fig. 3). Pairwise comparison disclosed an average of 20.7 (or 25%) amino acids differences (range, 3–34 or 4–40%).

Paur-DRB haplotype composition

The number and position of *DRB* loci are unknown in most non-human primates, but the *DRB* region typically involves a variable number of loci in other primates and usually a variety of alleles per locus. Baboons possessed two to eight (mean ± SEM, 5.36±1.6) different sequences, indicating several duplications of *DRB* genes in this species. Analysis of sequences present in mothers and offspring allowed identification of combinations of sequences that are inherited together, i.e., haplotypes (Table 1), further confirmed by linkage disequilibrium. A total of 11 distinct haplotypes, characterized by a combination of one to four *Paur-DRB* sequences, were described in 194 out of 199 individuals (Table 2). The five remaining individuals were heterozygous (i.e., carry two different *Paur-DRB* haplotypes) with one identified haplotype plus other sequences. These other sequences belong to haplotypes that could not be fully characterized as these combinations of sequences were only present in one individual.

Evidence for positive selection within *Paur-DRB* sequences

PSS were identified in *Paur-DRB* sequences using a maximum-likelihood analysis. The significant deviation of the LRT statistic from a χ^2 distribution (X^2=61.62, df=2, P<0.001) allows rejection of the null model assuming neutral evolution (M7) in favor of a model allowing for a class of sites subjected to diversifying selection (M8). Ten sites were detected as PSS by this approach, and nine were statistically significant (P<0.05) according to Bayes empirical analysis, which represents 11% of the amino acid sequence. The non-significant PSS was excluded from subsequent analyses. Eight PSS were identical to the ABS defined by homology with HLA (Brown et al. 1993; Figs. 3 and 4). The remaining PSS (amino acid position 77) was situated within a distance of two amino acids of an ABS. Eight human ABS (positions 2, 19, 21, 23, 24, 27, 66, and 79) were not identified as PSS in *P. ursinus*. The number of observed non-synonymous mutations only partially predicts the occurrence of PSS at a given site (Fig. 4) because this approach incorporates realistic codon substitution models as well as phylogeny of the sequences. Therefore, codon sites with numerous non-synonymous substitutions may not be detected as PSS (e.g., site 21) because the number of evolutionary steps, such as nucleotide substitutions, leading to this site configuration is not significantly greater than for any other site. Conversely, a site exhibiting two non-synonymous mutations (e.g., site 77) may be identified as

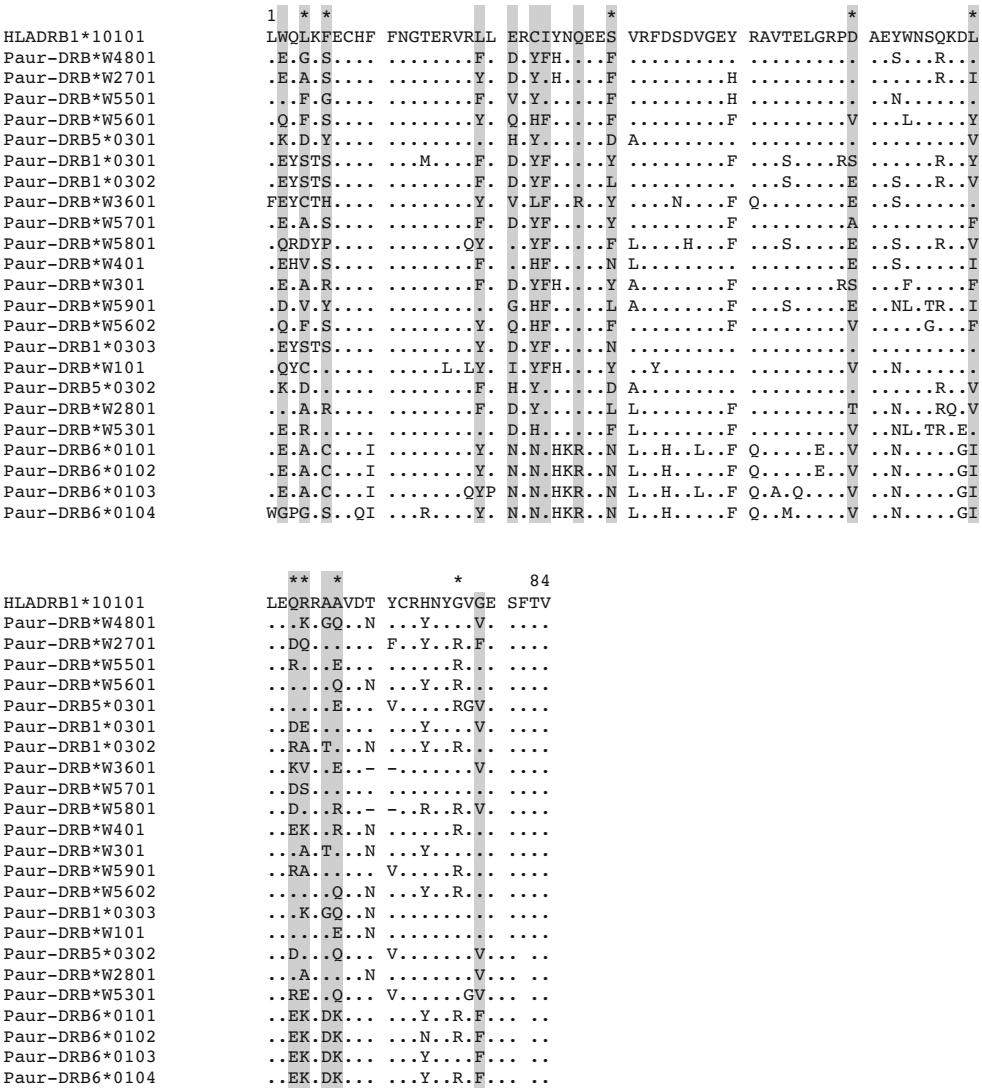

Fig. 3 Predicted protein sequences of chacma baboon *DRB* sequences (exon 2). Amino acid number is shown above the *HLA-DRB1*010101* allele. Identity to this allele is shown by *dots*, while differences are given by *letter substitutions* and gaps by *dashes*. Antigen-binding sites defined by homology with HLA are shown in *gray*. Chacma baboon positively selected sites (PSS) are designed by *stars*

PSS because the site configuration involves several independent evolutionary events. We cannot exclude a false-positive for this site, although Wong et al. (2004) demonstrates, using simulations on HLA data, that the maximum likelihood method has power and accuracy in detecting positive selection over a wide range of parameter values.

Of the 16 ABS defined by homology with HLA (Brown et al. 1993), all were variable. Thirty-five of 68 (51.5%) non-ABS were polymorphic. These variable non-ABS were mostly located next to an ABS. d_S and d_N were calculated separately for ABS and non-ABS. ω was significantly greater than unity in the ABS (Table 3). For comparative purposes, d_S and d_N were also calculated for PSS and non-PSS, and exhibited a similar pattern as ABS/non-ABS (Table 3). Non-synonymous substitutions exhibited by the four *DRB6* lineage sequences do not seem to be preferentially located within variable sites (Fig. 4). This result suggests that *Paur-DRB6* sequences may not be under the same selective pressures as other sequences.

Definition of supertypes

Using these nine significant PSS for each of the 23 sequences, 12 supertypes were defined by hierarchical clustering based on physicochemical amino acid properties (Fig. 2a). Such clustering method gives distinct results from clustering based on the number of nucleotide similarities (Fig. 2b) or on the number of amino acid similarities (Fig. 2c), as only three groups of sequences consistently cluster together: the four *Paur-DRB6* sequences, the two *Paur-DRB5* sequences, and the two sequences belonging to the *Paur-DRB*W56* lineage.

Evidence for intragenic recombination (or homologous gene conversion)

From the 23 *Paur-DRB* sequences, the population recombination rate (ρ) was estimated as ρ=52. This value is greater than the corresponding population mutation rate (θ),

Table 2 *Papio ursinus Mhc-DRB* haplotypes

Haplotype (*N* copies observed)	*DRB1* lineage	*DRB5* lineage	*DRB6* lineage	Other lineages
A (59)			*Paur-DRB6*0104* (S8)	*Paur-DRB*W4801* (S4)
				*Paur-DRB*W2701* (S12)
				*Paur-DRB*W5501* (S5)
B (101)		*Paur-DRB5*0301* (S7)		*Paur-DRB*W5601* (S5)
C (4)	*Paur-DRB1*0301* (S1)			
	*Paur-DRB1*0302* (S3)			
D (44)	*Paur-DRB1*0301* (S1)			*Paur-DRB*W5901* (S3)
E (82)	*Paur-DRB1*0302* (S3)		*Paur-DRB6*0101* (S8)	*Paur-DRB*W401* (S11)
				*Paur-DRB*W301* (S2)
F (41)			*Paur-DRB6*0102* (S8)	*Paur-DRB*W3601* (S9)
				*Paur-DRB*W5701* (S1)
G (21)	*Paur-DRB1*0303* (S4)			*Paur-DRB*W101* (S6)
				*Paur-DRB*W5701* (S1)
				*Paur-DRB*W5801* (S10)
H (30)		*Paur-DRB5*0301* (S7)		*Paur-DRB*W5602* (S5)
I (3)	*Paur-DRB1*0301* (S1)	*Paur-DRB5*0302* (S7)		
J (2)	*Paur-DRB1*0303* (S4)			*Paur-DRB*W101* (S6)
				*Paur-DRB*W2801* (S2)
K (5)				*Paur-DRB*W5301* (S6)
L[a] (1)	*Paur-DRB1*0302* (S3)			*Paur-DRB*W101* (S6)
	*Paur-DRB1*0303* (S4)			
M[a] (1)			*Paur-DRB6*0104* (S8)	*Paur-DRB*W4801* (S4)
				*Paur-DRB*W5701* (S1)
				*Paur-DRB*W5801* (S10)
N[a] (1)	*Paur-DRB1*0302* (S3)		*Paur-DRB6*0103* (S8)	*Paur-DRB*W5701* (S1)
	*Paur-DRB1*0303* (S4)			*Paur-DRB*W101* (S6)
O[a] (1)			*Paur-DRB6*0104* (S8)	*Paur-DRB*W4801* (S4)
				*Paur-DRB*W101* (S6)
				*Paur-DRB*W5801* (S10)

In parentheses, the corresponding supertype of each *Mhc-DRB* sequence
[a] These combinations of sequences were retrieved in only one individual and are therefore not fully characterized.

which was estimated as $\theta=29.3$. The ratio $\rho/\theta=1.8$ indicates an important contribution from recombination compared to mutation in the history of these sequences. The likelihood permutation test shows that the ρ estimate significantly differs from the value expected under the null hypothesis of no recombination ($P=0.01$).

Non-random association of sequences within haplotypes

The presence of two *DRB* sequences belonging to the same supertype was never observed within a haplotype. It is unlikely that this observation occurred by chance given that the observed value lies outside of 95% of the values generated under the hypothesis of a random distribution of supertypes within haplotypes (Table 4, Fig. 5a). The mean number of distinct supertypes per individual, which can be considered an efficient individual *Mhc-DRB* repertoire, is 4.2 in our population. This observed value is greater than 95% of the simulated values (Fig. 5b). The mean size of efficient individual repertoires calculated from simulated values was 3.7. The non-random allocation of supertypes within haplotypes thus increases the efficiency of an individual's *Mhc-DRB* repertoire by 13%, compared to neutral expectations. When grouping sequences on the basis of sequence similarity (Fig. 2b and c), we found that sequences were randomly associated within haplotypes, which means that two closely related sequences (proximity based either on amino acid or nucleotide sequences) can belong to a same haplotype (Table 4).

Discussion

Twenty-three distinct MHC class II *DRB* exon 2 sequences were detected in 199 chacma baboons (*P. ursinus*). This represents the first study of MHC variability in a natural population of Old World Monkeys, which makes interspecific comparisons difficult. Sequence diversity found in two natural populations of prosimians, fat-tailed dwarf lemurs (*Cheirogaleus medius*; Schwensow et al. 2007) and mouse lemurs (*Microcebus murinus*; Schad et al. 2004), exhibited 50 and 14 *Mhc-DRB* sequences, respectively, for compara-

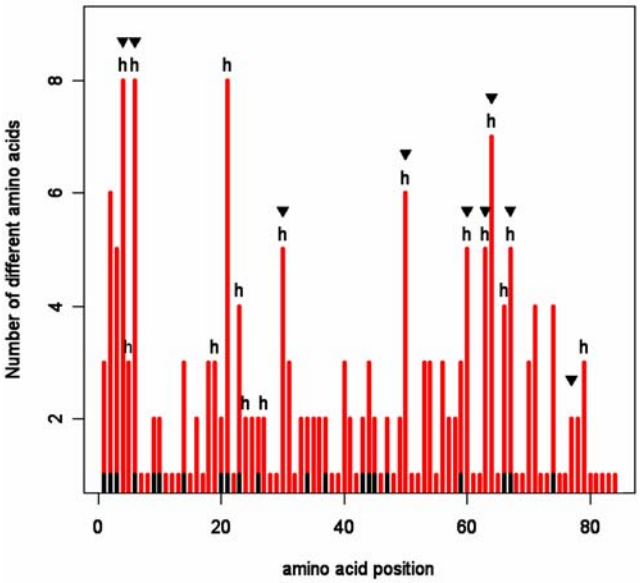

Fig. 4 Amino acid variability plot for 23 *Paur-DRB* sequences. The *horizontal axis* gives the amino acid position, whereas the *vertical axis* shows the number of different amino acids at a given position. *Letter* "h" indicate ABS defined by homology with HLA. Chacma baboon PSS are indicated by *black triangles*. The superimposed *black bars* represent the contribution of four *Paur-DRB* sequences from DRB6 lineage to the total amino acid variability. See text for details

ble sample sizes. A total of 126 *Mhc-DRB* sequences have been reported in rhesus macaques (*M. mulatta*), a more closely related species that is maintained in large captive colonies (around 1,000 animals from several origins; Otting et al. 2000). In comparison to these previous primate studies and taking into account the fact that the extensive polymorphism described in rhesus macaques is partly due to their various origins, our baboon population exhibits relatively limited but actually comparable levels of polymorphism for the *Mhc-DRB* region.

All nucleotide sequences had a distinct amino acid sequence, and there was no stop codons suggesting that any sequence was a pseudogene. However, we found several sequences assigned to the *DRB6* lineage, which is a pseudogene in both humans and rhesus macaques (Bontrop et al. 1999). Only a gene expression study could thus ensure that *Paur-DRB* sequences are all functional, and our results need to be interpreted cautiously until the means to study RNA are made more feasible at our field site in the Namib desert.

Each individual possessed between two and eight sequences, indicating up to four loci in the *Paur-DRB* region. One haplotype (haplotype K) appears to have only one *DRB* sequence. This may suggest that some *DRB* alleles have been missed during DNA amplification. However, we used a relatively low annealing temperature to ensure amplification of all *DRB* sequences in every individual (Huchard et al. 2006). We successfully amplified four *Paur-DRB6* sequences (that are particularly divergent from other *Paur-DRB* sequences) and recorded a high repeatability when analyzing PCR products from independent DNA amplifications originating from a same individual. Thus, we believe that our protocols were optimized to minimize the risk of selectively amplifying some *DRB* alleles.

The *DRB* region organization is highly variable in primates, representing different evolutionary strategies. Thirty-one genome configurations are reported for the *Mamu-DRB*-region of rhesus macaques with numbers of loci varying from two to eight, and where the *DRB1* locus can be absent, present, or even duplicated. This species is characterized by unstable haplotypes, subject to frequent gene shuffling, recombination-like events, and limited allelic polymorphism in comparison to HLA (Doxiadis et al. 2001; Doxiadis et al. 2000; Doxiadis et al. 2003). In contrast, five stable major-*DRB* region configurations are assigned to humans, with a number of loci differing from one to four and a *DRB1* locus in every configuration. Assuming homology between the *–DRB* lineages identified in baboons and the corresponding human *–DRB* loci, *Paur-DRB* haplotypes display up to ten different configurations from 11 ascertained haplotypes. Some haplotypes possess two *DRB1* loci, whereas others do not exhibit any. This suggests that some duplications are not fixed in the population. Moreover, the baboon haplotypes that could not be fully characterized suggest that some haplotypes may originate from recent recombination events, because they often share more than one *DRB* sequence with a

Table 3 The estimated rates (±SE) of non-synonymous (d_N) and synonymous (d_S) substitutions for antigen (ABS) or non-antigen (non-ABS) binding sites defined by homology with HLA, for codon sites identified as positively selected sites (PSS) and for codon sites identified as non-positively selected sites (non-PSS) within this set of *Paur-DRB* sequences, and including all sites in 23 *Paur-DRB*-exon 2 sequences

Positions	*n*	d_N	d_S	d_N/d_S	Z *(P)*
ABS	16	0.613±0.080	0.195±0.068	3.14	3.584 (0.000)
Non-ABS	68	0.081±0.015	0.116±0.029	0.69	1.050 (0.296)
PSS	9	0.840±0.067	0.176±0.065	4.77	6.648 (0.000)
Non-PSS	75	0.103±0.016	0.122±0.030	0.84	0.600 (0.550)
All	84	0.156±0.024	0.127±0.027	1.22	0.853 (0.395)

n the number of codons in each category, *P* the probability that d_N and d_S are not different using a *Z* test of selection

Table 4 Resampling test results

Grouping criterion	Number of functionally overlapping loci within haplotypes			Number of distinct supertypes per individual		
	Observed mean	Simulated mean	*P* value [CI]	Observed mean	Simulated mean	*P*-value [CI]
Nucleotide sequence similarity	5	3.82	0.839 [0.832–0.846]	3.78	3.59	0.286 [0.279–0.293]
Amino acid sequence similarity	5	5.83	0.448 [0.438–0.458]	3.52	3.17	0.175 [0.168–0.183]
Supertypes	0	3.38	0.022 [0.019–0.025]	4.23	3.68	0.009 [0.006–0.011]

Observed and simulated mean, *P*-value [95% Confidence Interval] is given for each test (see text for details on the tests)

known confirmed haplotype. *Paur-DRB* haplotypes thus exhibit a clear potential for generating new gene combinations, through locus deletion, duplication, or gene shuffling, despite a strong preferential *Mhc-DRB* linkage disequilibrium. Therefore, the limited allelic variation together with the number of configurations exhibited by the *Paur-DRB* region, which is clearly higher than in *HLA-DRB*, strongly recall the variable nature of *Mamu-DRB* organization.

To measure the occurrence of diversifying selection in MHC sequences, the majority of studies analyze ω in putative ABS defined by homology with HLA sequences. Recently, alternative approaches based on maximum-likelihood models have proven efficient in detecting PSS (Yang and Bielawski 2000; Yang et al. 2000) in species other than humans, including natural populations: mole-rats (*Heliophobius argenteocinereus*, *Heterocephalus glaber*, *Cryptomys hottentotus hottentotus*, *Cryptomys damarensis*; Kundu and Faulkes 2004), salmon (*Salmo salar*; Consuegra et al. 2005), chamois (*Rupicapra* spp; Schaschl et al. 2005), and fat-tailed dwarf lemurs (*Cheirogaleus medius*; Schwensow et al. 2007). In line with our results, these studies find a high but partial coincidence between PSS identified in a species and putative human ABS (Consuegra et al. 2005; Schaschl et al. 2005; Schwensow et al. 2007). Moreover, the higher estimation of ω in PSS than in ABS suggests that maximum-likelihood methods identifying species-specific PSS are presumably more accurate and less sensitive to false-positive risks than methods relying on homology with HLA sequences to detect functionally important codon sites in species other than humans.

Supertype classification in our baboon population was based on 23 sequences and cannot be validated using experimental antigen-binding assays. *Paur-DRB* sequences were clustered within 12 supertypes based on physico-chemical properties of PSS following a method recently applied in a free-ranging primate, the fat-tailed dwarf lemur, who found 11 different *Mhc-DRB* supertypes (Schwensow et al. 2007). The definition of supertypes based on sequence motifs is clearly an imperfect or at least incomplete representation of binding, but this classification is certainly a meaningful approximation of peptide specificity supported by recent human and non-human primate studies (Lund et al. 2004; Schwensow et al. 2007; Sette and Sidney 1998; Southwood et al. 1998).

The presence of two *DRB* sequences belonging to the same supertype was never observed within a haplotype. Further analyses provide statistical support to this observation, which is unlikely to occur by chance. Moreover, control analyses using clustering based on amino-acid or nucleotide sequence similarity show that highly similar

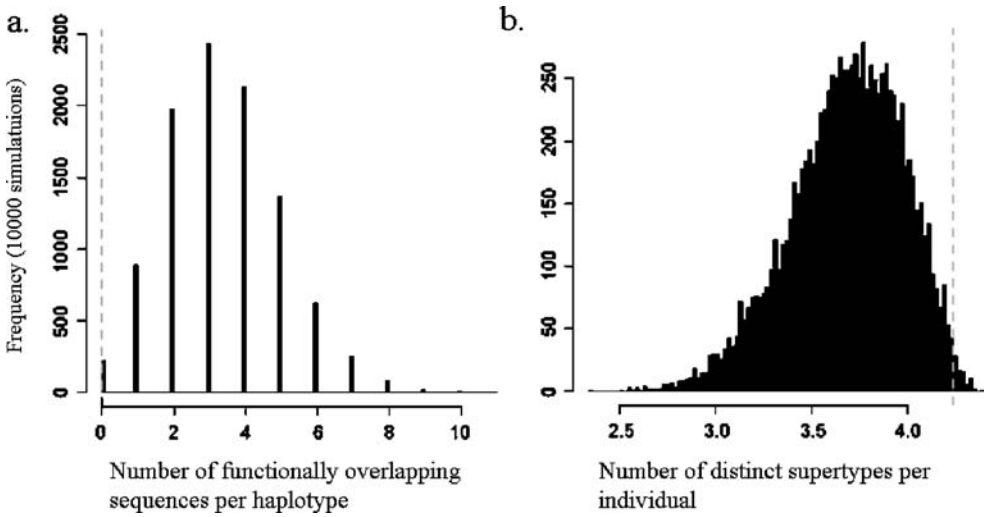

Fig. 5 Support for non-random association of MHC supertypes within *Mhc-DRB* haplotypes, excluding unconfirmed haplotypes. The *dotted lines* indicate the position of the observed value. **a** Frequencies of the number of functionally overlapping sequences (i.e., from a same supertype) within a same haplotype, generated under a random distribution. **b** Frequencies of the number of distinct supertypes per individual generated under a random distribution. See text for details

sequences can be associated within a same haplotype. The non-random association of supertypes within haplotypes is thus unlikely to result from the fact that two sequences grouped into the same supertype are alleles with shared histories (i.e., from a same locus) that cannot belong to a same haplotype. The discordance observed between the hierarchical clustering tree based on the physico-chemical properties of PSS and the neighbor-joining tree built using the number of differences between *Paur-DRB* amino acid sequences confirms that two sequences grouped within a supertype do not necessarily share the high level of sequence similarity characterizing sequences of a same lineage (except for the *DRB6*, which was excluded from simulation tests, and two sequences from the *DRB5* lineage). Thus, *Mhc-DRB* haplotype composition seems to be optimized in order to avoid association of functionally redundant sequences. Co-expression of duplicated loci might be selectively favored in the MHC to increase the diversity of parasite species recognized (Nuismer and Otto 2004). Similarly, this non-random association of supertypes within haplotypes that significantly increases the individual number of supertypes possessed (or individual efficient repertoire) may increase the range of parasites recognized (i.e., heterozygote advantage, Doherty and Zinkernagel 1975).

Different patterns exhibited by the HLA-*DRB* region on the one hand, and the *Paur*- and *Mamu-DRB* regions on the other hand suggest different evolutionary strategies in these species (Doxiadis et al. 2001). While the human population invested mainly in generating a high degree of allelic variation at the various *DRB* loci, the rhesus macaque and chacma baboon populations primarily generated a large number of singular combinations of *DRB* loci (de Groot et al. 2004). In line with recent MHC studies (Richman et al. 2003a; Richman et al. 2003b; Schaschl et al. 2005), the estimated population recombination rate (ρ) exceeds the estimated mutation rate (θ) in baboons. This indicates that the generation of new recombinant alleles exceeds that of alleles derived by new point mutations. This high rate of recombination (or homologous gene conversion), which is further confirmed by the relative instability of *Paur*- and *Mamu*- haplotypes, probably results in a constant turnover of allelic combinations through a birth and death type process. Such variability will favor the emergence and subsequent selection of advantageous combinations. The relatively limited allelic variation in these monkey species may increase the selective pressure for generating efficient allelic combinations representing the best possible coverage of the pathogen repertoire. Thus, a functional tuning of haplotypes in the *Mhc-DR* region may be part of the evolutionary pathway followed by macaques and baboons. Testing such a hypothesis in macaques or other close relatives, preferentially in natural populations, would represent a further step in clarifying the evolutionary mechanisms underlying the interspecific differences in the organization of the *Mhc-DR* region of primates.

Acknowledgment We would like to thank Arnaud Berthomieu and Jo Osborn for technical assistance in the lab and Alexandre Courtiol for help in the resampling tests. We thank the Ministry of Lands and Resettlement for permission to work at Tsaobis Leopard Park, the Desert Research Foundation of Namibia for affiliation, and the Ministry of Environment and Tourism for research permission. We confirm that we have adhered to the Guidelines for the Use of Animals in Behavioural Research and Teaching (*Animal Behaviour*, 65:249–255, 2003) and the legal requirements of the country (Namibia) in which the work was carried out. This work was funded by a Natural Environment Research Council (UK) grant awarded to G. C. and a Ministère de l'Education et de la Recherche (France) Studentship awarded to E.H. G.C. was funded by a NERC Advanced Fellowship during the writing of this paper. This paper is a publication of the ZSL Institute of Zoology's Tsaobis Baboon Project. Contribution ISEM 2008-054.

References

Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, Yunis EJ (2006) The haplotype structure of the human major histocompatibility complex. Hum Immunol 67:73–84 doi:10.1016/j.humimm.2005.11.006

Anisimova M, Nielsen R, Yang ZH (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Apanius V, Penn D, Slev P, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. Crit Rev Immunol 17:179–224

Bodmer WF (1972) Evolutionary significance of the HL-A system. Nature 237:139–183 doi:10.1038/237139a0

Bontrop RE, Otting N, de Groot NG, Doxiadis GGM (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339–350 doi:10.1111/j.1600-065X.1999.tb01403.x

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 364:33–39 doi:10.1038/364033a0

Consuegra S, Megens HJ, Schaschl H, Leon K, Stet RJM, Jordan WC (2005) Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Mol Biol Evol 22:1095–1106 doi:10.1093/molbev/msi096

Cowlishaw G (1999) Ecological and social determinants of spacing behaviour in desert baboon groups. Behav Ecol Sociobiol 45:67–77 doi:10.1007/s002650050540

de Groot N, Doxiadis GG, de Groot NG, Otting N, Heijmans C, Rouweler AJM et al (2004) Genetic makeup of the DR regions in rhesus macaques: gene content, transripts, and pseudogenes. J Immunol 172:6152–6157

Delguercio MF, Sidney J, Hermanson G, Perez C, Grey HM, Kubo RT et al (1995) Binding of a peptide antigen to multiple Hla alleles allows definition of an A2-like supertype. J Immunol 154:685–693

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52 doi:10.1038/256050a0

Doxiadis GGM, Otting N, de Groot N, Noort MC, Bontrop RE (2000) Unprecedented polymorphism of *Mhc-DRB* region configurations in rhesus macaques. J Immunol 164:3193–3199

Doxiadis GGM, Otting N, de Groot N, Bontrop RE (2001) Differential evolutionary MHC class II strategies in humans and rhesus macaques: relevance for biomedical studies. Immunol Rev 183:77–85 doi:10.1034/j.1600-065x.2001.1830106.x

Doxiadis GGM, Otting N, de Groot NG, de Groot N, Rouweler AJM, Noort R et al (2003) Evolutionary stability of MHC class II haplotypes in diverse rhesus macaque populations. Immunogenetics 55:540–551 doi:10.1007/s00251-003-0590-9

Doytchinova IA, Flower DR (2005) In silico identification of supertypes for class II MHCs. J Immunol 174:7085–7095

Doytchinova IA, Guan PP, Flower DR (2004) Identifiying human MHC supertypes using bioinformatic methods. J Immunol 172:4314–4323

Gu X, Nei M (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Mol Biol Evol 16:147–156

Hedrick PW, Kim KJ (2000) Genetics of the complex polymorphisms: parasites and maintenance of the major histocompatibility complex variation. In: Singh RS, Krimbas CB (eds) Evolutionary genetics: from molecules to morphology. Cambridge University Press, Cambridge, pp 204–234

Huchard E, Cowlishaw G, Raymond M, Weill M, Knapp LA (2006) Molecular study of Mhc-DRB in wild chacma baboons reveals high variability and evidence for trans-species inheritance. Immunogenetics 58:805–816 doi:10.1007/s00251-006-0156-8

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. Annu Rev Genet 32:415–432 doi:10.1146/annurev.genet.32.1.415

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? Microbes Infect 2:1335–1341 doi:10.1016/S1286-4579(00)01287-9

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic, New York, pp 21–132

Khazand M, Peiberg C, Nagy M, Sauermann U (1999) *Mhc-DQ-DRB* haplotype analysis in the rhesus macaque: evidence for a number of different haplotypes displaying a low allelic polymorphism. Tissue Antigens 54:615–624 doi:10.1034/j.1399-0039.1999.540612.x

Knapp LA, Cadavid LF, Eberle ME, Knechtle SJ, Bontrop RE, Watkins DI (1997) Identification of new Mamu-DRB alleles using DGGE and direct sequencing. Immunogenetics 45:171–179 doi:10.1007/s002510050186

Kumar S, Koichiro T, Jakobsen IB Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics Application Note 17:1244–1245

Kundu S, Faulkes CG (2004) Patterns of MHC selection in African mole-rats, family Bathyergidae: the effects of sociality and habitat. Proc R Soc Lond B Biol Sci 271:273–278 doi:10.1098/rspb.2003.2584

Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P et al (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 55:797–810 doi:10.1007/s00251-004-0647-4

Madden DR (1995) The 3-dimensional structure of peptide-Mhc complexes. Annu Rev Immunol 13:587–622 doi:10.1146/annurev.iy.13.040195.003103

Marsh SGE (2005) Nomenclature for factors of the HLA system, update June 2005. Tissue Antigens 66:338–340 doi:10.1111/j.1399-0039.2005.00479.x

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

Myers RM, Fischer SG, Lerman LS, Maniatis T (1987) Nearly all single base substitutions in DNA fragments joined to GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res 13:3131–3145 doi:10.1093/nar/13.9.3131

Nei M, Gojobori T (1986) Simple methods for estimating the number of synonymous and non-synonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A 94:7799–7806 doi:10.1073/pnas.94.15.7799

Nei M, Hughes AL (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) Eleventh histocompatibility workshop and conference. Oxford University Press, Oxford, England, pp 27–38

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford

Nuismer SL, Otto SP (2004) Host–parasite interactions and the evolution of ploidy. Proc Natl Acad Sci U S A 101:11036–11039 doi:10.1073/pnas.0403151101

Otting N, de Groot NG, Noort MC, Doxiadis GGM, Bontrop RE (2000) Allelic diversity of *Mhc-DRB* alleles in rhesus macaques. Tissue Antigens 56:58–68 doi:10.1034/j.1399-0039.2000.560108.x

Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC Class I genes: divergence time estimates. Mol Biol Evol 20:601–609 doi:10.1093/molbev/msg064

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 *Peromyseus maniculatus*. Genet Res 82:89–99 doi:10.1017/S0016672303006347

Robinson J, Waller MJ, Parham P, de Groot N, Bontrop RE, 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 doi:10.1093/nar/gkg070

Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 41:2481–2491 doi:10.1021/jm9700575

Schad J, Sommer S, Ganzhorn JU (2004) MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. Conserv Genet 5:299–309 doi:10.1023/B:COGE.0000031137.50239.d3

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 doi:10.1007/s00251-005-0784-4

Schwensow N, Fietz J, Dausmann KH, Sommer S (2007) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. Heredity 99:265–277 doi:10.1038/sj.hdy.6800993

Sette A, Sidney J (1998) HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr Opin Immunol 10:478–482 doi:10.1016/S0952-7915(98)80124-6

Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50:201–212 doi:10.1007/s002510050594

Sette A, Livingston B, McKinney D, Appella E, Fikes J, Sidney J et al (2001) The development of multi-epitope vaccines: Epitope identification, vaccine design and clinical evaluation. Biologicals 29:271–276 doi:10.1006/biol.2001.0297

Sette A, Newman M, Livingston B, McKinney D, Sidney J, Ishioka G et al (2002) Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. Tissue Antigens 59:443–451 doi:10.1034/j.1399-0039.2002.590601.x

Seutin G, White BN, Boag PT (1990) Preservation of avian blood and tissue samples for DNA analysis. Can J Zool 69:82–90 doi:10.1139/z91-013

Sidney J, Howard MG, Kubo RT, Sette A (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. Immunol Today 17:261–266 doi:10.1016/0167-5699(96)80542-1

Slierendregt BL, Otting N, van Besouw N, Jonker M, Bontrop RE (1994) Expansion and contraction of rhesus macaque DRB regions by duplication and deletion. J Immunol 152:2298–2307

Slierendregt BL, van Noort JT, Bakas RM, Otting N, Jonker M, Bontrop RE (1992) Evolutionary stability of transpecies major histocompatibility complex class II DRB lineages in humans and rhesus monkeys. Hum Immunol 35:29–39 doi:10.1016/0198-8859(92)90092-2

Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S et al (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. J Immunol 160:3363–3373

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL et al (1994) Crystal-structure of the human class-Ii Mhc protein Hla-Dr1 complexed with an influenza-virus peptide. Nature 368:215–221 doi:10.1038/368215a0

Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data. Bioinformatics 18:207–208 doi:10.1093/bioinformatics/18.1.207

Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E et al (2003) Advantage of rare HLA supertype in HIV disease progression. Nat Med 9:928–935 doi:10.1038/nm893

Wong WSW, Yang ZH, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051 doi:10.1534/genetics.104.031153

Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503 doi:10.1016/S0169-5347(00)01994-7

Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449