Immunogenetics (2006) 58: 191–202
DOI 10.1007/s00251-006-0085-6

ORIGINAL PAPER

J. Bryja · M. Galan · N. Charbonnel · J. F. Cosson

Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the *DQA* MHC class II gene in voles (Arvicolinae)

Received: 22 July 2005 / Accepted: 5 January 2006 / Published online: 9 February 2006

Abstract Major histocompatibility complex (MHC) genes play important role in host–parasite interactions and parasites are crucial factors influencing the population dynamics of hosts. We described the structure and diversity of exon 2 of the MHC class II *DQA* gene in three species of voles (Arvicolinae) exhibiting regular multi-annual fluctuations of population density and analysed the processes leading to the observed MHC polymorphism. By using cloning–sequencing methodology and capillary electrophoresis-single strand conformation polymorphism, we described seven sequences in the water, eight in the common, and seven in the bank voles coming from an area of 70 km^2 around the Nozeroy canton in the Jura Mountains (Franche Comté, France). All exon 2 sequences translate to give unique amino acid sequences and positive selection was found to act very intensively on antigen binding sites. We documented the presence of recombination at vole *DQA* region but its importance in generating allelic polymorphism seems to be relatively limited. For the first time within rodents, we documented the duplication of the *DQA* gene in all three species with both copies being transcriptionally active. Phylogenetic analysis of allelic sequences revealed extensive trans-species polymorphism within the subfamily although no alleles were shared between species in our data set. We discuss possible role of parasites in forming the recent polymorphism pattern of the *DQA* locus in voles.

J. Bryja · M. Galan · N. Charbonnel · J. F. Cosson
Centre de Biologie et Gestion des Populations (UMR 22),
INRA, Campus International de Baillarguet,
CS 30016, 34988 Montferrier sur
Lez Cedex, France

J. Bryja (✉)
Department of Population Biology,
Institute of Vertebrate Biology,
Academy of Sciences of the Czech Republic,
67502 Studenec 122, Czech Republic
e-mail: bryja@brno.cas.cz
Tel.: +420-568-627950
Fax: +420-568-627950

Keywords Muridae · *Arvicola* · *Microtus* · *Clethrionomys* · Allelic diversity

Introduction

The major histocompatibility complex (MHC) is a family of highly polymorphic genes occurring in vertebrates and encoding a set of transmembrane proteins. The function of these proteins is to present peptides to specific receptors of T cells of the immune system, thus triggering various immune responses (Klein 1986; Roitt 1997). The hallmark of MHC genes is the extremely high degree of polymorphism, the extent of which being the highest among all vertebrate protein-coding loci (Klein 1986). There are several lines of evidence that polymorphic MHC loci are subject to balancing selection (see review in Hughes and Yeager 1998a). In particular, MHC polymorphism is characterized by a trans-species character mainly observed for MHC class II loci. Allelic lineages are maintained over long periods of time, even across speciation events (Klein 1987; Figueroa et al. 1988). Trans-species sharing of MHC sequences among genera has been supported by observations in fish (Ottova et al. 2005 and references therein), rodents (Edwards et al. 1997; Figueroa et al. 1988; Musolf et al. 2004; Seddon and Baverstock 2000), carnivores (Hedrick et al. 2000a; Seddon and Ellegren 2002), ungulates (Hedrick et al. 2000b) and primates (Otting et al. 2002). The pressure exerted by pathogens is considered the major component of balancing selection influencing MHC polymorphism. None of the other selection pressures proposed for the maintenance of MHC polymorphism, i.e. kin recognition, inbreeding avoidance and mate choice (see Apanius et al. (1997) for an extensive review) have the ubiquitous distribution among the vertebrate taxa that the parasite-targeted immune response does (Klein and O'Huigin 1994).

Most of the MHC studies are primarily interested in the extent of polymorphism within loci. In the MHC, however, there is substantial amount of polymorphism due to gene duplication, insertion or deletion. In mammals, the MHC genes are supposed to evolve by "birth and death" model

(Nei et al. 1997). Little evidence exists that interlocus gene conversion plays an important role in the evolution of mammal MHC (Hughes and Yeager 1998b). Within each class II region, loci have sometimes duplicated independently in different mammalian groups, and half the loci are functional (Hughes and Nei 1990). Gene duplication plays a major role in the adaptive evolution of organisms (Hughes 1994). Exploring the evolutionary history of MHC and understanding the redundancy of MHC genes is, thus, very important in evolutionary and ecological studies using them as genetic markers.

Muridae is the largest mammalian family (Musser and Carleton 1993). The complete MHC class II region was sequenced in rats (Günther and Walter 2001; Hurt et al. 2004) and mice (Blake et al. 2003). The most genetic variation was assured by homologues of the human DQ (called RT1.B in rat, H2-A in mouse) and DR (RT1.D in rat, H2-E in mouse) genes. While genes coding for α chain of DQ and DR molecules and that for β chain of DQ molecule occur only in single copy, the gene coding for β chain of DR is duplicated in both species (but the second copy is probably pseudogene, at least, in rats; Easterfield et al. 2003). In other rodent species studied till today, the structure of DQ and DR regions seem to be similar with a single copy of the DQA (Pfau et al. 1999; Seddon and Baverstock 1999; Smulders et al. 2003; Sommer et al. 2002) and DQB (T. Hambuch, unpublished data), and one (Musolf et al. 2004; Richman et al. 2002; Smulders et al. 2003; Sommer et al. 2002) or two (Sommer et al. 2002) copies of the DRB gene. The only exception is the genus *Peromyscus*, where more than one DQB genes were found (Richman et al. 2002). Evidence of trans-species persistence of allelic lineages in murids is still ambiguous. Although the ancestral polymorphism of MHC class II genes was supported by some authors (Edwards et al. 1997; Figueroa et al. 1988; Musolf et al. 2004; Seddon and Baverstock 2000), others did not found support for trans-species persistence of MHC alleles among different murid taxa (Pfau et al. 1999; Richman et al. 2003a).

In this study, we analysed MHC variation in the Arvicolinae subfamily as it has not been considered in the previous analyses although it is a relevant model to provide insight into the MHC polymorphism. The evolutionary history of voles is well known through molecular phylogeny (Jaarola et al. 2004; Martin et al. 2000). Moreover, many species of voles exhibit regular fluctuations in population density (Butet and Spitz 2001). MHC genes are likely to be involved in these demographic fluctuations through their potential influence on reproduction (e.g. in humans Gill 1992; Sauermann et al. 2001) and immunity, with host–pathogen interactions being a possible cause of vole cycles (e.g. Cavanagh et al. 2004). The objectives of this study were (1) to describe the structure of MHC-*DQA* diversity in the water vole *Arvicola terrestris*, which exhibits multi-annual regular cycles in the East of France (Berthier et al. 2005), and (2) to reconstruct the evolution of MHC-*DQA* loci; first, by examining allelic diversity and duplication events in three species of voles and, second, by comparison with homologous sequences of other rodents.

Materials and methods

Animals

We analysed the MHC class II gene *DQA* in three species of voles coming from the Jura Mountains (Franche Comté, France). The study area covers about 70 km^2 around the Nozeroy canton (46°48′ N, 6°02′ E). Common voles (*Microtus arvalis*) and bank voles (*Clethrionomys glareolus*) were collected in two close localities while water voles (*A. terrestris*) came from five different localities covering the whole area. The animals were live-trapped using INRA traps for *M. arvalis* and *C. glareolus* and BTS traps for *A. terrestris*. The end of the tail or one finger was fixed in 95% ethanol as material for DNA extraction. Approximately 10 individuals per species were killed by cervical dislocation and the spleen was placed in TRIzol (Invitrogen) or RNAlater (Ambion) for analyses of MHC expression on RNA level.

DNA extraction and PCR

Genomic DNA was extracted from tail or finger tissues by DNeasy Tissue Kit (Qiagen). DNA was extracted according to manufacturer's instructions and finally eluted by 400 µl of AE buffer. We amplified the complete exon 2 of the *DQA* (*RT1.Ba*) homologue of the MHC class II. The primers RT1.Ba 390C and RT1.Ba 619NC were taken from Seddon and Baverstock (1998) and the amplifications (marked as PCR 1) were performed in the following conditions: 2.5 mM $MgCl_2$, 0.1 µM of each primer, 100 µM dNTPs, 0.1 U Taq polymerase (Qiagen) in appropriate 1× polymerase chain reaction (PCR) buffer and 1 µl of extracted DNA. Deionised water was added to a 10-µl reaction volume. The thermal profile started with initial denaturation at 94°C (2 min) followed by 30 cycles of denaturation at 92°C (10 s), annealing at 52°C (10 s) and extension at 72°C (1 min) and final extension at 72°C for 10 min. The PCR amplifications were performed in the PTC-100 PCR thermocycler (MJ Research). Every amplification reaction gave a single band of approximately 280 bp. By cloning and sequencing of cDNA (see below), we found that used primer pair did not sufficiently amplify two additional sequences; more specifically, these two sequences were amplified only in the absence of other sequences. Therefore, the second PCR (i.e. PCR 2) was performed. In this PCR, we amplified a specific *DQA* band of approximately 300 bp only in animals bearing one or both of these two sequences (no product was detected for alleles well amplified in PCR 1) with modified forward primer (DQAintron1-F: 5′-CCACCATCCACCTAAATT CCTCAG-3′) and reverse primer RT1.Ba 619NC (Seddon and Baverstock 1998) by using almost the same protocol as in PCR 1 (the only difference was annealing for 30 s at 50°C). This amplification also gave a non-specific band of approximately 340 bp that served as a control of successful amplification. Thus, the genotyping of the exon 2 of the *DQA* gene was based on two PCR amplifications and

subsequent two analyses of single strand conformation polymorphism (SSCP) in a capillary.

Capillary electrophoresis-single strand conformation polymorphism

The exon 2 of the *DQA* locus was amplified in two PCRs by using fluorescently labelled primers (forward by 6′-FAM and reverse by HEX) and protocols described above. We checked the successful PCR by electrophoresis on ethidium-bromide-stained agarose gel and then performed single strand conformation polymorphism analysis of PCR products by capillary electrophoresis (CE) on MegaBACE 1000 DNA Analysis System (Amersham Biosciences), which was equipped with 96 coated capillaries (for details, see Bryja et al. 2005). In brief, PCR products were 20 times diluted by TrisCl, pH 8.5, mixed with 0.25 μl of ROX-labelled internal marker MegaBACE ET 400-R Size Standard (Amersham Biosciences) to a final volume of 10 μl, denatured by 2 min in 95°C in PCR machine and, after snap-cooling on ice, injected to the capillaries. We used 3% MegaBACE non-denaturing Long Read Matrix (Amersham Biosciences) as a sieving matrix and MegaBACE 1× Running Buffer (Amersham Biosciences) supplied with 10% glycerol was used as a running buffer. The raw data were exported in text format and different runs were aligned and peaks were detected in the software QUISCA (Higasa et al. 2002).

Cloning and sequencing

We selected 10 water voles with diverse SSCP patterns to obtain the sequences of the *DQA* exon 2. The *DQA* gene was amplified as described above, but using non-labelled primers. The PCR products were purified by the QIAquick PCR purification kit (Qiagen), ligated to vectors and transformed to bacteria using pGEM-T Vector System II (Promega). Positive transformants containing insert of approximate length were identified by PCR screening and agarose electrophoresis. We subsequently isolated 10 clones for each presumable heterozygote (more than one peak in at least one colour in CE-SSCP pattern) and eight clones for each presumable homozygote (one peak in both FAM and HEX in CE-SSCP). The plasmid DNA was extracted by QIAprep Spin Miniprep Kit (Qiagen) and insert was sequenced by using DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) and primer SP6. Sequences were obtained on MegaBACE 1000 DNA Analysis System (Amersham Biosciences).

cDNA synthesis and analysis

Total RNA was extracted from TRIzol (Invitrogen) or RNAlater (Ambion) preserved samples (six water, four common and four bank voles) by using TRIzol/chloroform extraction. RNA was then precipitated by isopropanol, washed by 75% ethanol, and resuspended in 20 μl RNase-free water. We used 1 μl of extracted RNA for reverse transcription in a 20-μl reaction using the following conditions: 1 μl oligo$(dT)_{18}$ (500 μg/ml), 1 μl dNTPs (10 mM), 4 μl RT5× buffer (Invitrogen), 2 μl DTT (0.1 M), 1 μl RNase OUT (40 U/μl) and 1 μl M-MLV reverse transcriptase (Invitrogen). Total cDNA was diluted 100 times before subsequent specific PCR amplification. Complete exon 2 of the *DQA* gene was amplified using primers At-cDNA-DQAex2-F (5′-TGTGGAGGTCAA GACGACAT-3′) and At-cDNA-DQAex2-395-R (5′-AAAGCAGATGAGGGTGTTGG-3′) designed on the basis of homology with rat cDNA. The primer sequences are located within exon 1 or exon 3, respectively, thus avoiding possible amplification of contaminating genomic DNA. PCR amplification was performed in a 50-μl volume using the following conditions: 0.1 μM of each primer, 1× PCR buffer, 2.5 mM $MgCl_2$, 0.1 mM dNTPs and 0.1 U of Taq polymerase (Qiagen). The PCR products were purified, cloned and sequenced in the same way as described above.

Sequence analysis

The sequences were edited and aligned in BioEdit Sequence Alignment Editor (Hall 1999) using ClustalW Multiple Alignment. Previous analysis yielded that relatively high proportion of obtained sequences could be artefacts of PCR amplifications (Bryja et al. 2005). We considered a new sequence variant as a new allele only when it met the criteria summarised by Kennedy et al. (2002). The criteria are, that when using DNA cloning and sequencing, there have to be at least three identical clones, identified in either two separate PCRs from the same individual or from PCRs from at least two different individuals. Nomenclature of the water vole (Arte-*DQA*), common vole (Miar-*DQA*) and bank vole (Clgl-*DQA*) sequences follows Klein et al. (1990). The sequences of complete exon 2 of three vole species are available in GenBank under the following accession numbers: DQ137805-DQ137826.

The published sequences of the entire exon 2 of the MHC Class II gene *DQA* for six other rodents of the Muridae family were used for comparative and phylogenetic analyses (used abbreviations and GeneBank accession numbers in parentheses): five Murinae, *Rattus tunneyi tunneyi* (Rtt, AF041070), *Rattus leucopus cooktownensis* (Rlc, AF145094), *Rattus villosissimus* (Rv, AF145096), *Rattus norvegicus* (Rano, AJ554214, L11340, M22366, X14879) and *Mus musculus* (Mumu, K01924 and K01926) and one Sigmodontinae, *Peromyscus leucopus* (Pele, U34805). Human *Homo sapiens* (Human M17236), sheep *Ovis aries* (OvAr1, M33304; OvAr2, M33305) and pig *Sus scrofa* (PigM29938) sequences were used as the outgroups in phylogenetic analysis. Moreover, phylogenetic analyses were also performed on shorter sequences (212 bp, nucleotide positions 20 to 232) to include data from another Sigmodontinae species, the cotton rat,

Sigmodon hispidus, as published by Pfau et al. (1999). As this did not change the results on voles, only analyses on sequences of complete exon 2 are presented thereafter.

Molecular evolutionary analyses were conducted using MEGA version 3.0 (Kumar et al. 2004) to calculate relative frequencies of synonymous (dS) and nonsynonymous (dN) substitution by the method of Nei and Gojobori (1986) with the Jukes and Cantor (1969) correction for multiple substitutions. Amino acid positions involved in peptide binding were identified by comparison with the crystal structure of murine class II molecules I-A^k (Fremont et al. 1998), I-A^d (Scott et al. 1998) and I-A^{g7} (Latek et al. 2000). The structure of described protein alleles slightly differs and, finally, we decided to take into account amino acids number 11, 22, 24, 32, 52, 53, 54, 55, 58, 61, 62, 65, 66, 68, 69, 72, 73 and 76, i.e. the residues that are referred as antigen binding sites (ABS) in at least two of the above studies (numbering of amino acid residues is according to Brown et al. 1993).

More rigorous analyses of selection were performed using likelihood ratio modelling in the program CodeML, which is contained in the PAML 3.14 program suite (Yang 1997). We used maximum likelihood models which account for the heterogeneity among site partitions by using different ω (=dN/dS) parameters for the partitions. We used the random-sites models (Yang et al. 2000) assuming several heterogeneous site classes with different ω parameter without a priori knowledge which class each site is from. The models implemented in this study are M0, M1a, M2a, M3, M7, and M8 (Yang et al. 2000, details on models M1a and M2a in Yang et al. 2005). The one-ratio model, M0, assumes a single ω ratio for all codon sites. Model M1a (NearlyNeutral) assumes two classes of sites in the protein: the conserved sites under purifying selection at which $0<\omega_0<1$ (and their proportion p_0) and selectively neutral sites at which $\omega_1=1$ (and their proportion p_1). Model M2a (PositiveSelection) adds a third class of sites with ω as free parameter, thus allowing for sites with $\omega>1$. Model M3 (discrete) uses a general discrete distribution with three site classes, with the proportions (p_0, p_1 and p_2) and the ω ratios (ω_0, ω_1 and ω_2) estimated from the data. Model M7 (β model) is based on the β distributions and allows the ω ratio to take values of between zero and unity. Finally, model M8 (β and ω) adds an extra class of sites to the M7, with the proportion and the ω ratio estimated from the data, thus allowing for sites with $\omega>1$. The likelihood ratio test statistics for comparing nested models are calculated as follows: $2(L_b-L_a)$ is compared with χ^2 distribution with P_b-P_a degrees of freedom, where L_a and L_b are log-likelihood values and P_a and P_b are the number of parameters for each of the models being compared. By the LRTs we compared M0 with M3, M1a with M2a, and M7 with M8, respectively. When the alternative models (M2a, M3 and M8) suggest the presence of sites with $\omega>1$, all three tests can be considered tests of positive selection (Yang et al. 2000). Posterior probabilities for site classes have been calculated by Bayes empirical Bayes (BEB) method in models M2a and M8 (Yang et al. 2005). If the posterior means of ω for some site classes are >1 (calculated as the average of ω over all site classes weighted by the posterior probabilities), those sites are likely to be under positive selection (Yang et al. 2005).

We used the program Geneconv 1.81 (Sawyer 1999) to find the most likely candidate fragments for recombination/gene conversion events between vole sequences of the *DQA* gene. This method has one of the highest probabilities of correctly inferring gene conversion when it is present (Posada 2002). Global and pairwise statistical tests were performed with 10,000 permutation runs. The software has an integrated Bonferroni correction that protects against the inflation of type I errors of the many pairwise tests that are performed. For example, when using the alignment of 22 sequences of 246 bp (i.e. all three species together), it results in 22×21/2=231 possible pairs of sequences. The corrected *p*-value (we used the Karlin and Altschul 1990 *p*-values) was thus multiplied by 231 to obtain a global α-value of 0.05, which is a very conservative and recommended procedure (Sawyer 1999). This method,

```
            [          1 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778 8888]
            [ 34567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234]
Arte-DQA*01   DHVGSYGI TTYQSYGPNG QFTFEFDGDE LLYVDLDKKE TVWRLPEFAQ LRRFDPQGGL QNIATGKHNL DILTKRSNFT PATN
Arte-DQA*02   ..I..... .V........ .Y.....D.. .......... ....I...G. .G.L...... R....A.... .....T.... ....
Arte-DQA*03   ..I..... .V........ .Y.....D.. .......... ....I...G. ...L...... R....A.... .V...T.... ....
Arte-DQA*04   ........ .......... .......... .......... .......... .......... .....E.... .V........ ....
Arte-DQA*05   ..I..... .VN..... .Y.....D.. .......... ....I...G. .G.L...... R....A.... .V...T.... ....
Arte-DQA*06   .....F.. DF..T.ES.. .Y........ EF....G... ....I...G. .TT....... RE..SS.Y.. ..MI....S. ....
Arte-DQA*07   ..I..F.. IVC....S.. .Y.M...... .F........ ....I...GH .AS....I.. .E..GA.Y.. .LM.....S. QV..
Miar-DQA*01   ..I..... NV.....S.. .Y.M...... E.....G... ....I.G.G. ..S..... ....VA.... .MWI.E.... ....
Miar-DQA*02   ..I..... NV.....S.. .Y.M...... KF....G... ....I...GK .TS..... .E..RA.F.. ...I.ET.S. .T..
Miar-DQA*03   ..I..... EL........ .Y........ .F....G... ....I...GH .AS..... .EM..A.... .V......S. ....
Miar-DQA*04   .....F.T VV........ .YIM...... .......... ....I...GK .ASL...S.. .Q...A.... .V...M..S. .V..
Miar-DQA*05   ....TF.T VV........ .V.C...... .......... ....I...GK .ASL..... .Q...A.F.. .V......S. .V..
Miar-DQA*06   ....I... .......... .......... E......... ...K....S. ...L...... R....M.Y.. ........S. L...
Miar-DQA*07   N..... .......... .......... E......... ....I...G. .ITV...... .......... .V........ ....
Miar-DQA*08   ....T... .......S.. ...M..N... E.....G... ....I...G. ..SY...... ....VE.F.. .L........ ....
Clgl-DQA*01   ..IA.... .......... .......... K.......R. ....I...D. .IT..... ....S..F.. .V.I.Q..S. ....
Clgl-DQA*02   ....A... IVC......D .Y.M...... .F...S.... ....I...G. .AS..... .Q...A.... ..VI....S. .T..
Clgl-DQA*03   .....F.. DF.....S.. .Y........ EF...F.... ....I...G. .TS..... .E..LS.QD. ..MI....S. ....
Clgl-DQA*04   .......L SFS..HQ... .Y........ .F...S.... ....I...G. .TS..... .S..IE.... .NMI.W..S. ....
Clgl-DQA*05   N.I..... IV........ .Y........ EF........ ....I...G. .LY....N.. HD..IV.QY. .P......S. ....
Clgl-DQA*06   ....T... .......S.. ...M...... E.....G... ....I...G. ..SY...... ....VE.F.. .L........ ....
Clgl-DQA*07   ....T... .......S.. ...M...... E.....G... .I..I...G. ..S..... ....A..F.. .L........ ....
```

Fig. 1 Alignment of seven expressed amino acid sequences of the complete *DQA* exon 2 of the water voles (Arte-*DQA*), eight sequences of the common voles (Miar-*DQA*) and seven sequences of the bank voles (Clgl-*DQA*). *Dots* indicate identity in the amino acid sequence of the Arte-*DQA**01. Numbering of amino acid residues is according to Brown et al. (1993)

Table 1 Composition of the Arte-*DQA* exon 2 genotypes in 96 water voles on the basis of the capillary electrophoresis-single strand conformation polymorphism

Sequence						No. of individuals
01 or 04[a]	02	03	05	06	07	
x					x	1
x	x					4
x (1×01)	x			x		3
x (1×01)	x	x				6
x (2×04)	x		x			11
x (2×04)		x	x			8
x (1×01, 1×04)		x				3
x		x		x		4
x		x			x	3
x (1×04)			x			10
x			x	x		13
x			x		x	6
	x					3
	x	x				2
	x			x		4
	x				x	2
		x				1
		x		x		3
		x			x	1
			x	x		1
				x	x	7
Total						96

[a]Sequences Arte-*DQA**01 and Arte-*DQA**04 are indistinguishable by CE-SSCP analysis. Numbers in parentheses indicate the number of individuals with particular sequence identified by the cloning and sequencing

however, may have low power in the presence of the extensive recombination detected at some MHC loci. Thus, we estimated the amount of recombination rate (ρ=4$N_e r$) in the history of a set of aligned sequences by Hudson's (2001) composite-likelihood estimate approach using the programme LDhat (McVean et al. 2002). This programme estimates recombination rates fairly accurately with regard to sequences evolving under symmetric balancing selection (Richman et al. 2003b). The value of the composite likelihood recombination (ρ_{LCE}) is estimated conditioned on the estimate of a mutation rate per site (θ=4$N_e\mu$), obtained using a finite-series version of the Watterson estimator. The value of ρ_{LCE} was compared statistically to the hypothesis of no recombination (ρ=0) using a permutation test implemented in LDhat. Finally, the distribution of the correlation between r^2 measure of linkage disequilibrium and physical distance was calculated under random permutation of the physical position of the single nucleotide polymorphisms (1,000 permutations) and compared with correlation obtained from the data. This analysis, implemented also in LDhat, was used as another test for the presence of recombination (Awadalla and Charlesworth 1999; McVean et al. 2002).

Paup version 4.0b10 for Microsoft Windows 95/NT was employed to construct a phylogenetic tree of the *DQA* sequences based on 247 bp of the exon 2 (base pairs no. 3–249 of the complete exon), applying the neighbour-joining (NJ) algorithm with Kimura's two-parameter distances. A bootstrap analysis (1,000 replicates) was performed to determine the reliability of the branching.

Results

Identification of allelic diversity in water voles

By using cloning–sequencing methodology and CE-SSCP analysis, we analysed 78 clones derived from genomic DNA from 10 water voles and identified seven different sequences of the complete Arte-*DQA* exon 2 (Fig. 1). All sequences were confirmed by comparing CE-SSCP patterns of cloned PCR products with those of the individual from which the sequence was isolated. All sequences are distinguishable by CE-SSCP with one exception, i.e. Arte-*DQA**01 and Arte-*DQA**04 that produced the same CE-SSCP pattern even if we used various electrophoresis conditions (temperature, sieving matrix, etc.). We have not found additional *DQA* sequences by CE-SSCP analysis of another 86 water voles from the same localities.

Table 2 Synonymous and non-synonymous distances for antigen-binding sites (ABS) and non-ABS in *DQA* sequences

	N	ABS			Non-ABS		
		dS	dN	dN/dS	dS	dN	dN/dS
All rodents	32	0.28±0.10	0.39±0.07	1.40	0.30±0.06	0.09±0.02	0.31
Murinae (5 spp.)	9	0.22±0.16	0.21±0.05	0.95	0.16±0.05	0.08±0.02	0.53
Arvicolinae (3 spp.)	22	0.21±0.08	0.43±0.07	2.02	0.20±0.04	0.07±0.02	0.36
Clethrionomys	7	0.28±0.16	0.49±0.09	1.74	0.20±0.05	0.08±0.02	0.37
Microtus	7	0.19±0.09	0.45±0.10	2.40	0.18±0.05	0.08±0.02	0.42
Arvicola	8	0.11±0.07	0.36±0.07	3.16	0.17±0.05	0.06±0.01	0.33

Synonymous (dS) and non-synonymous (dN) distances were calculated using the Nei–Gojobori method with a Jukes–Cantor correction. Standard errors were estimated by bootstrap with 1,000 replicates
N Number of analysed sequences

Table 3 Summary of test statistics for the likelihood-ratio tests of exon 2 of the *DQA* gene in three species of voles

Models compared	*df*	Test statistic	Significance
M1a vs M2a	2	19.794	$p<0.001$
M3 vs M0	4	82.860	$p<0.001$
M8 vs M7	2	22.665	$p<0.001$

Test statistic was computed as $2(L_b–L_a)$, where L_a and L_b are log-likelihood values for each of the nested models being compared

Expression pattern

In total, we sequenced cDNA fragments of 50 clones from six water voles with diverse CE-SSCP pattern (identified previously in 24 water voles). This analysis revealed seven different sequences of transcribed exon 2 of the DQA gene (Fig. 1). These sequences were identical to those characterised on the basis of genomic DNA suggesting that there are no pseudogenes of the *DQA* gene or at least that pseudogenes are not amplified by the used primers in water voles.

We performed similar analysis of expressed gene diversity also in two other vole species. After the CE-SSCP analysis of 16 common and 16 bank voles, the cloning and sequencing of cDNAs from four common voles with diverse CE-SSCP pattern (43 clones were sequenced) revealed eight different sequences, while in the same number of bank voles (38 clones), seven sequences were transcribed (Fig. 1).

Number of DQA genes in voles

In six water voles (from 16 individuals used to assess the *DQA* sequence diversity on genomic and cDNA level), three sequences were identified by both the SSCP and cloning–sequencing analyses. In total, 96 individuals were genotyped by CE-SSCP, and 54 individuals (56%) exhibited three peaks in at least one colour in their CE-SSCP pattern. These results evidence occurrence of at least three different sequences that are amplified by used primers, i.e. at least two copies of the *DQA* gene exist in the water voles (Table 1). All individuals with three *DQA* sequences exhibited CE-SSCP peak corresponding to alleles Arte-*DQA**01 and Arte-*DQA**04 (they are not distinguishable by CE-SSCP—see above) and we have never found individuals with four CE-SSCP sequences. These observations suggest that these two alleles could be located on the second *DQA* locus. Considering this fact, both loci were amplified in 72 (75%) water voles, while only one locus (bearing alleles Arte-*DQA**02, 03, 05, 06 and 07) was present in 24 individuals (Table 1).

Based on the cloning and sequencing of cDNAs, the *DQA* gene is duplicated also in another two vole species. In three common voles (from four analysed), we found three transcribed sequences; the fourth individual expressed only two sequences. In four bank voles analysed for cDNA variation, we similarly found one to four expressed sequences, suggesting that, at least in some individuals, the *DQA* gene is duplicated. Additional PCR amplification (only PCR 1 was performed) and CE-SSCP of *DQA* gene on genomic level revealed three or four peaks in one colour in seven from 16 common voles and three peaks (four peaks have never been observed) in six from 16 bank voles.

Sequence analysis

In 22 sequences found within three vole species, 130 out of 247 nucleotide positions were variable, among which 80 were parsimony informative. No indels were detected. The mean divergence in the *DQA* exon 2 sequences ranged from 10.9 to 14.6% nucleotide substitutions within the three vole species. These values indicate the divergence both within and between loci as we were not able to distinguish alleles of particular loci. Taking into account the fact that MHC polymorphism might be expressed not only by true locus allelism, in further analyses, we considered all sequences as if they would be alleles of one locus.

Table 4 Results of maximum likelihood models of exon 2 of the *DQA* gene in three species of voles

Model code	*P*	Log-likelihood	Parameter estimates	Positively selected sites
M0 (one ratio)	1	−1764.296	ω=0.861	None
M1a (NearlyNeutral)	1	−1701.588	p_0=0.551, p_1=0.450	Not allowed
M2a (PositiveSelection)	3	−1681.795	p_0=0.502, p_1=0.304, p_2=0.195, ω_2=3.312	*24*, *31*, **52**, *53*, **65**, **66**, **68**, **72**
M3 (discrete)	5	−1681.436	p_0=0.542, p_1=0.292, p_2=0.167, ω_0=0.110, ω_1=1.297, ω_2=3.653	Not analysed
M7 (beta)	2	−1705.451	p=0.149, q=0.151	Not allowed
M8 (beta and omega)	4	−1682.785	p_0=0.792, p_1=0.208, p=0.211, q=0.300, ω=3.231	*11*, **24**, **31**, **52**, **53**, *62*, **65**, **66**, **68**, **72**, *76*

Alignment of 22 sequences from three vole species was used as input for the program CodeML (included in the PAML 3.14 program suite). Positively selected sites were identified in models M2a and M8 by the Bayes empirical Bayes procedure (Yang et al. 2005). Sites inferred under selection at the 99% level are listed in bold, and those at the 95% level are in italic

P Number of parameters in the ω distribution, ω selection parameter, p_n proportion of sites that fall into the ω_n site class, p, q shape parameters of the β function (for models M7 and M8)

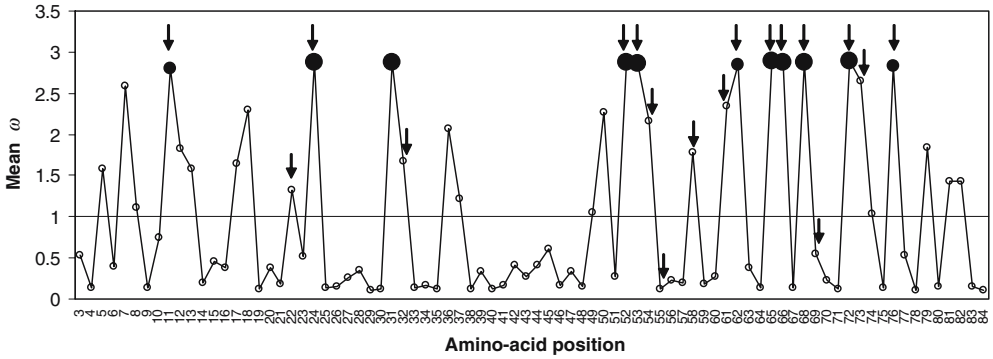

Fig. 2 Posterior means of ω, calculated as the average of ω over the 11 site classes, weighted by the posterior probabilities under the random-sites model M8 (β and ω). The posterior probabilities were computed by the Bayes empirical Bayes procedure in the program CodeML implemented in the PAML3.14 package. The amino acid numbering is according to Brown et al. (1993). *Arrows* indicate the ABS positions identified on the basis of the crystal structure of the I-A proteins (details in the text). Sites inferred to be under positive selection at the 99% level are indicated by *large black circles* and those at the 95% level by the *medium black circles*

All exon 2 sequences identified in voles translate to give unique amino acid sequences (Fig. 1), reflecting the high level of non-synonymous substitutions at these loci. We observed strong and systematic increases in ratio of non-synonymous vs synonymous substitutions (dN/dS) at the ABS as expected under balancing selection (Table 2). This was especially true within the Arvicolinae subfamily.

We applied the maximum-likelihood-based random-sites models to 246 bp of the *DQA* exon 2 encoding amino acids 3–84 (numbering is according to Brown et al. 1993) using PAML3.14. The results are summarised in Tables 3 and 4. The models allowing for the inclusion of positive selection, i.e. M2a, M3 and M8, fitted data significantly better than simpler models M1a, M0 and M7 (Table 3). The random-sites models, thus, demonstrate extreme variability in selective pressure among sites in the exon 2 and the presence of a number of sites under positive selection. These sites were identified by using the Bayes empirical Bayes procedure incorporated in PAML3.14. In total, eight sites were significantly positively selected under the model M2a and 11 sites under the M8 (Table 4). In model M8, posterior means of ω were >1 at the next 21 sites, suggesting the presence of other positively selected sites (similar results obtained under model M2a are not shown). The positively selected sites almost exactly correspond to amino acid residues lying in the ABS (Fig. 2).

Using the program Geneconv 1.81, we did not found strong evidence for intra- or interlocus recombination in the alignment of 22 sequences of 246 bp (base pairs no. 3–248) of three species of voles. Global tests did not identify gene conversion, which is defined as the presence of a global inner fragment. However, 18 from 231 pairwise comparisons were significant at $p<0.05$ and, even if they became insignificant after the Bonferroni correction for multiple comparisons, this number is higher than predicted by chance (~11=5% from 231) indicating the presence of gene conversion in our sample. The ambiguous inference on the presence of recombination was obtained also on the basis of analyses in the program LDhat (Table 5). The amount of recombination was higher than that of mutation in the bank and common voles, while lower in the water voles. Permutation tests, however, significantly rejected the null hypothesis ($\rho=0$) and confirmed the presence of recombination in water and common voles. Negative correlation of linkage disequilibrium and physical distance was significant only in water voles. Both measures of recombination were significant in the alignment of sequences of all three species suggesting the presence of recombination events in the common history of arvicoline rodents (Table 5).

Phylogenetic analysis resulted in a single shortest tree (Fig. 3). The sequences of the five Murinae species formed a clade well distinct (bootstrap value of 67) from the ones of the Sigmodontinae and Arvicolinae species. By contrast, the 22 sequences from the three Arvicolinae species do not form monophyletic groups respective to the species level. This result suggests trans-species polymorphism within that subfamily although no alleles were shared between species in our data set. Based on the phylogenetic analysis,

Table 5 Estimates of the amount of population mutation (Waterson's $\theta=4N\mu$) and population recombination rate ($\rho=4Nr$) obtained using the programme LDhat (McVean et al. 2002)

Species	n	θ	ρ	p(ρ)	r^2	p(r^2)
A. terrestris	55	23	3	**0.019**	−0.078	**0.015**
C. glareolus	61	25	35	0.071	−0.021	0.166
M. arvalis	68	26	99	**0.017**	−0.037	0.056
All three species	88	24	88	**0.001**	−0.034	**0.029**

Significance values of ρ and r^2 are based on 1,000 permutations and values of $p<0.05$ are in bold
n Number of segregating sites, r^2 correlation of pairwise linkage disequilibrium with nucleotide distance

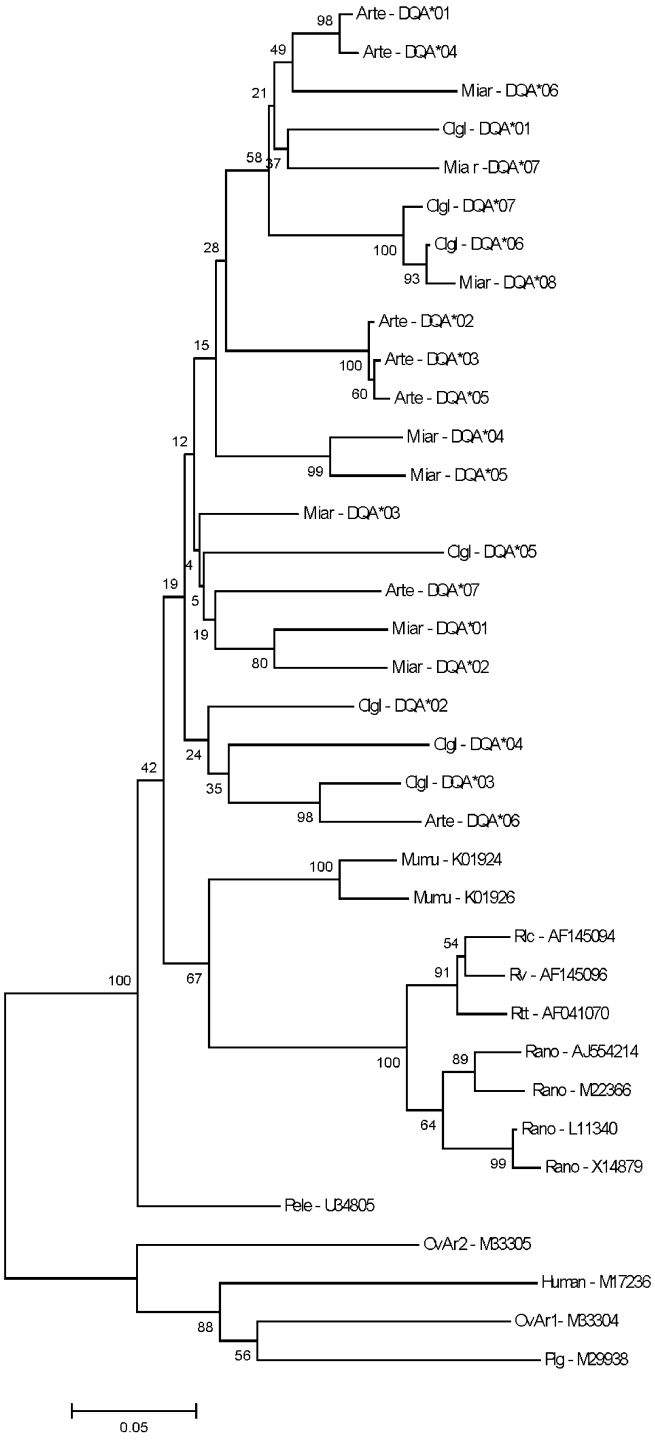

Fig. 3 Phylogenetic relationships of the MHC class II gene *DQA* sequences in nine species of rodents expressed as the neighbour-joining tree based on Kimura's two-parameter distances. Bootstrap values based on 1,000 replicates are shown. GenBank accession numbers from previously described *DQA* sequences follow the species designation. Arte *A. terrestris*, Miar *M. arvalis*, Clgl *C. glareolus*, Rano *Rattus norvegicus*, Rlc *Rattus leucopus cooktownensis*, Rtt *Rattus tunneyitunneyi*, Rv *Rattus villosissimus*, Mumu *Mus musculus*, Pele *Peromyscus leucopus*. *DQA* exon 2 sequences of human, sheep *O. aries* (*OvAr*, both *DQA1* and *DQA2* genes) and pig were used as outgroups

it is not possible to distinguish alleles from different loci, but we suggest that duplication predated speciation event.

Discussion

In the present study, we described for the first time the diversity of the MHC class II *DQA* gene in voles. We identified seven sequences in the water voles, eight in the common voles and seven in the bank voles. The *DQA* locus, hence, seems to be quite variable considering the facts that analysed animals originated from the area of about only 100 km^2 and the low numbers of analysed common and bank voles. Similar studies on rodents indicated even higher allelic richness at the *DQA* gene (13–36 alleles); however, they were usually focussed on the population genetic structure using MHC markers, i.e. they analysed more animals from larger areas (Seddon and Baverstock 1999; Pfau et al. 2001; Gouy de Bellocq et al. 2005). This suggests that the number of existing sequences in voles will increase after more individuals from large area will be typed.

Commonly used measure of selection, i.e. dN/dS test, demonstrated that there is strong evidence for an increased rate of substitution at non-synonymous vs synonymous sites in the ABS within *DQA* exon 2 in voles. The *DQA* locus in voles, thus, seems to be subject to a strong balancing selection. This result is in accordance with published data on *DQA* gene in rodents, where the dN/dS ratio is usually much higher at ABS sites than at non-ABS ones, even if the differences are only rarely significant (see review in Pfau et al. 1999). Different results are rare at the exon 2 of the MHC Class II gene and are usually considered as unexpected (Gouy de Bellocq et al. 2005) and/or reflecting saturation of the limited number of non-synonymous sites at the ABS (Richman et al. 2003b). The ambiguity of results could be caused by the unknown detailed structure of MHC proteins in the majority of species studied and, hence, incorrectly identified ABS sites. The ABS sites have been usually identified on the basis of the three-dimensional structure of human HLA-DR1 protein (Brown et al. 1993) which can be different from DQ and DR molecules in other taxa. The structure of the ABS (i.e. the position of amino acid residues incorporated in the peptide binding) is variable even between different alleles at one locus (for a review, see differences between that of Fremont et al. 1998; Latek et al. 2000; and Scott et al. 1998). Structural studies permit the identification of sites potentially involved in antigen binding, but do not expect all of them to be under balancing selection. The conserved sites in ABS then decrease the power of dN/dS analysis. From this point of view, it seems very useful to identify the sites under various types of selection by using maximum likelihood models. We presented extreme variability in selective pressure among sites in the exon 2 and the presence of a number of sites under balancing selection. In accordance with structural models, the majority of these sites have been previously known to be implemented in the antigen binding. However, we found additional sites under positive

selection not previously mentioned in the ABS (7, 18, 36, 50 and especially 31 that was not mentioned to be incorporated in antigen-binding in I-A^d and I-A^k alleles; Fremont et al. 1998; Scott et al. 1998). On the other hand, amino acid residues at positions 22 and 69 referred from murine ABS seem to be conserved or neutral to selection.

We clearly documented that at least two transcribed copies of the *DQA* gene occur in three species of voles. The duplication event predated the divergence of vole genera, which means that the common ancestor of all three vole genera before 7.4–9.3 Mya (Martin et al. 2000; Michaux and Catzeflis 2000; but see 5.5–6.6 Mya in Steppan et al. 2004) already had two copies of the *DQA* genes. Voles are, hence, the first group of rodents where the duplication of the *DQA* gene is documented. By sequencing the complete MHC region, it was found that the *DQA* homologue occurs in a single copy in rat (named *RT1.Ba*; Easterfield et al. 2003; Günther and Walter 2001; Hurt et al. 2004) and mouse (named *H2-Aa*; Blake et al. 2003) genome. A maximum of two alleles per individual corresponding to one copy of the *DQA* gene were found in the cotton rats *S. hispidus* (Pfau et al. 1999, 2001), three species of Malagasy rodents (Sommer and Tichy 1999; Sommer et al. 2002), the common hamster *Cricetus cricetus* (Smulders et al. 2003) and several species of the Australian rats (Seddon and Baverstock 1999, 2000). Nevertheless, the duplication of the *DQA* gene was documented outside the rodents. Two copies (one is transcribed, the second—called *DQA2*—is pseudogene) exist in human MHC (The MHC Sequencing Consortium 1999) and in other hominoid species of primates (non-human primate MHC database; Robinson et al. 2005). Two functional and highly polymorphic *DQA* genes were described in sheep (Fabb et al. 1993; Snibson et al. 1998 and references therein) and even a higher degree of complexity was found in the cattle *DQ* subregion, where three copies of the gene could occur (Ballingall et al. 1997).

Based on the CE-SSCP pattern, we tried to divide the *DQA* sequences in water voles to particular loci, and we suppose that both copies of the *DQA* gene are not always present in the vole MHC. More specifically, 25% of water voles had in their genotypes only sequences from presumably one locus. The lack of sequences from the second locus could be also caused by the occurrence of null alleles; however, the probability of null allele occurrence was decreased by using different primer pairs at the genomic and cDNA levels, respectively. The number of *DQA* genes varying with haplotype has been so far described only in artiodactyls. Between 10 and 18% of sheep lack the *DQA1* gene (Fabb et al. 1993) and various combinations of three *DQA* genes per haplotype were described in cattle (Andersson and Rask 1988; Ballingall et al. 1997). Such complex structure seems to be more common at *DRB* loci and is well described in California sea lions (Bowen et al. 2004), rhesus macaques (Khazand et al. 1999) and human (Andersson et al. 1994).

There is considerable uncertainty about the importance of gene conversion (intra-locus recombination) in driving the allelic diversity of MHC genes (Martinsohn et al. 1999; Ohta 1999). Recent application of efficient statistical methods to DNA sequences in non-model mammals indicated that intra-locus recombination can play a more important role in the generation of allelic polymorphism at *DRB* locus than mutation does (Richman et al. 2003a,b; Schaschl et al. 2005). In this study, we analysed for the first time the presence of recombination at the *DQA* loci and found ambiguous results. The analyses in LDhat programme suggest recombination among the DQA sequences in the common history of arvicolid rodents (see similar results in two chamois species; Schaschl et al. 2005); however the influence of recombination on forming *DQA* variation in particular species is not convincing. The inability to detect recombination unambiguously could be caused by the low number of analysed sequences (they were obtained from a low number of individuals from a relatively small area) or by the presence of duplicated gene, as we were not able to identify alleles of particular loci. Inter-locus recombination is thought to play a minor role in MHC evolution of mammals (Hughes and Yeager 1998b; but is well documented in fish and birds—for a review, see Edwards et al. 1995; Reusch et al. 2004; Wittzell et al. 1999); however, its absence can significantly decrease the power of statistical analyses of recombination when analysed sequences come from two or more loci.

In our study, we documented the trans-species polymorphism, i.e. the non-neutral retention of MHC sequences across three genera of voles causing discordance between MHC sequences and species trees. Even if the trans-species polymorphism of MHC genes has been well evidenced in various groups of vertebrates (see references in Introduction), in rodents the results have been ambiguous. Ambiguity in trans-species polymorphism might depend on the investigated locus with varying selection pressures acting on different loci. The ancestral polymorphism of MHC class II genes within different murid species was evidenced on *DRB* (Edwards et al. 1997; Smulders et al. 2003; Musolf et al. 2004) and *DQA* locus (Seddon and Baverstock 2000); however, others did not found support for trans-species persistence of alleles at these loci (Vestberg et al. 1998; Pfau et al. 1999; Richman et al. 2003a). The majority of recent subfamilies of Muridae separated during the initial bushlike radiation before 17–20 Mya and the calculated divergence time for the three arvicoline genera analysed in this study is between 6 and 8 Mya (Martin et al. 2000; Michaux and Catzeflis 2000; Steppan et al. 2004). The upper limit for taxa to share alleles is estimated to be 30–40 million years (Takahata and Nei 1990). Hence, the trans-species polymorphism in voles may be explained by the retention of ancestral polymorphism predating the speciation of Arvicolinae. This supposes that the founding population was large enough relative to divergence time to prevent the stochastic losses of allelic lineages through genetic drift and bottleneck (Hughes et al. 1994) and that the selection pressure on this locus was strong enough to maintain polymorphism over a long period of time. Nevertheless, it should be noted that, in the presence of recombination (which is still ambiguous at *DQA* gene in voles), the phylogenetic analyses should be considered with caution as the length of terminal branches

and the total branch lengths are larger and the time to the most recent common ancestor is smaller than that of a phylogenetic tree reconstructed with no recombination (Schierup and Hein 2000).

Pathogens are considered to be a major factor influencing balancing selection and trans-species evolution on MHC as they may be common among vertebrate taxa whether kin-selection or maternal–foetal incompatibility may not (Klein and O'Huigin 1994). Klein and O'Huigin (1994) proposed that parasites whose evolution follows Farhrenholz's rule that common ancestors of present day parasites could be the parasites of the common ancestors of present day hosts are primarily responsible for the evolution and properties of MHC polymorphism. MHC class II genes develop immune response against extra cellular pathogens such as bacteria and helminths. Several recent studies indicated the functional significance of MHC class II genes in defence against helminths in humans (Godot et al. 2000; Quinnell 2003), lemurs (Schad et al. 2005), sheep (Paterson et al. 1998) and rodents (Harf and Sommer 2005; Meyer-Lucht and Sommer 2005). By comparing parasitological surveys conducted on Murinae and Arvicolinae (S. Morand, personal communication), we were able to select six different helminths that can be considered as potential candidates for the trans-species evolution of MHC polymorphism as they were only present in voles. Although three possible cestodes (*Anoplocephaloides dentate*, *Paranoplocephala omphalodes*, and *Anoplocephaloides macrocephala*) were recorded, *P. omphales* and *A. macrocephala* were not considered as good candidates because recent phylogenetic studies revealed that they are likely to form host-specific complex species (Haukisalmi and Henttonen 2003; Haukisalmi et al. 2004). One trematode (*Plagiorchis proximus*) and two nematodes (*Heligmosomoides glareoli* and *Trichuris arvicolinae*) were also only found in Arvicolinae. Further experiments should focus on the detection of associations between resistance/susceptibility to these macroparasites and *DQA* allelic lineages. Indeed, previous studies revealed that *DQA* alleles could be strongly associated with resistance to trematodes (*Schistosoma japonicum*) in humans (Waine et al. 1998) and with increased susceptibility to nematodes (*Trichostrongylus colubriformis*) in sheep (Hohenhaus and Outteridge 1995).

We demonstrated that duplicated *DQA* gene in voles (more specifically, the sites included in ABS) is under strong effect of balancing selection, i.e. the gene is probably highly functional in the immunological response. This is the first gene of MHC so far characterised in voles, i.e. in the group of species frequently used as models in population ecology due to the ability of their populations to fluctuate regularly and multi-annually in their density. One of the leading hypotheses explaining the maintenance of high diversity of MHC genes in populations is the hypothesis of pathogen-driven balancing selection (Hughes and Yeager 1998a) and the parasites are recently thought to play very important role in cyclic population dynamics of voles (Cavanagh et al. 2004). It will be, hence, very useful to analyse the interactions of fluctuating population density, occurrence of parasites and frequency of the MHC genotypes in vole populations.

Acknowledgements The authors acknowledge the assistance of Julie Deter and Yannick Chaval in the provision of samples, Serge Morand for access to his database on rodent parasites and three anonymous referees for useful comments on the manuscript. This work received financial support from the Institut National de la Recherche Agronomique, the Region of Franche-Comté, the DIREN Franche-Comté, and the Ministry of the Environment of France. This publication has been also partially funded under the EU 6th Framework Program (GOCE-CT-2003-010284 Emerging Diseases in a Changing European Environment (EDEN) and is officially catalogued by the EDEN Steering Committee as EDEN0006). Its content does not represent the official position of the European Commission and is entirely under the responsibility of the authors.

References

Andersson L, Rask L (1988) Characterization of the MHC class II region in cattle. The number of *DQ* genes varies between haplotypes. Immunogenetics 27:110–120

Andersson G, Andersson L, Larhammar D, Rask L, Sigurdardottis S (1994) Simplifying genetic locus assignment of HLA-*DRB* genes. Immunol Today 15:58–62

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

Awadalla P, Charlesworth D (1999) Recombination and selection at *Brassica* self-incompatibility loci. Genetics 152:413–425

Ballingall KT, Luyai A, McKeever DJ (1997) Analysis of genetic diversity at the *DQA* loci in African cattle: evidence for a BoLA-DQA3 locus. Immunogenetics 46:237–244

Berthier K, Galan M, Foltete JC, Charbonnel N, Cosson JF (2005) Genetic structure of the cyclic fossorial water vole (*Arvicola terrestris*): landscape and demographic influences. Mol Ecol 14:2861–2871

Blake JA, Richardson JE, Bult CJ, Kadin JA, Eppig JT, Mouse Genome Database Group (2003) MGD: the mouse genome database. Nucleic Acids Res 31:193–195

Bowen L, Aldridge BM, Gulland F, Van Bonn W, DeLong R, Melin S, Lowenstine LJ, Stott JL, Johnson ML (2004) Class II multiformity generated by variable MHC-*DRB* region configurations in the California sea lion (*Zalophus californianus*). Immunogenetics 56:12–27

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 64:33–39

Bryja J, Galan M, Charbonnel N, Cosson JF (2005) Analysis of major histocompatibility complex class II gene in water voles using capillary electrophoresis-single stranded conformation polymorphism. Mol Ecol Notes 5:173–176

Butet A, Spitz F (2001) Cyclic fluctuations of microtine populations: half a century of research. Rev Ecol (Terre Vie) 56: 353–372

Cavanagh RD, Lambin X, Ergon T, Bennett M, Graham IM, van Soolingen D, Begon M (2004) Disease dynamics in cyclic populations of field voles (*Microtus agrestis*): cowpox virus and vole tuberculosis (*Mycobacterium microti*). Proc R Soc Lond B 271:859–867

Easterfield AJ, Bradley JA, Bolton EM (2003) Complementary DNA sequences encoding the rat MHC class II RT1-B^{u} and RT1-D^{u} α and β chains. Immunogenetics 55:344–350

Edwards SV, Wakeland EK, Potts WK (1995) Contrasting histories of avian and mammalian MHC genes revealed by class II B sequences from songbirds. Proc Natl Acad Sci U S A 92:12200–12204

Edwards SV, Chesnut K, Satta Y, Wakeland EK (1997) Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. Genetics 146:655–668

Fabb SA, Maddox JF, Gogolin EKJ, Baker L, Wu MJ, Brandon MR (1993) Isolation, characterization and evolution of ovine major histocompatibility complex class II DRA and *DQA* genes. Anim Genet 24:249–255

Figueroa F, Gunther E, Klein J (1988) MHC polymorphism predating speciation. Nature 355:265–267

Fremont DH, Monnaie D, Nelson CA, Hendrickson WA, Unanue ER (1998) Crystal structure of I-A^k in complex with a dominant epitope of lysozyme. Immunity 8:305–317

Gill TJ (1992) Influence of MHC and MHC-linked genes on reproduction. Am J Hum Genet 50:1–5

Godot V, Harraga S, Beurton I, Tiberghien P, Sarciron E, Gottstein B, Vuitton DA (2000) Resistance/susceptibility to *Echinococcus multilocularis* infection and cytokine profile in humans. II. Influence of the HLA *B8*, *DR3*, *DQ2* haplotype. Clin Exp Immunol 121:491–498

Gouy de Bellocq J, Delarbre C, Gachelin G, Morand S (2005) Allelic diversity at the Mhc-*DQA* locus of woodmouse populations (*Apodemus sylvaticus*) present in the islands and mainland of the northern Mediterranean. Glob Ecol Biogeogr 14:115–122

Günther E, Walter L (2001) The major histocompatibility complex of the rat (*Rattus norvegicus*). Immunogenetics 53:520–542

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Harf R, Sommer S (2005) Association between major histocompatibility complex class II *DRB* alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. Mol Ecol 14:85–91

Haukisalmi V, Henttonen H (2003) What is *Paranoplocephala macrocephala* (Douthitt, 1915) (Cestoda: Anoplocephalidae)? Syst Parasitol 54:53–69

Haukisalmi V, Wixkström LM, Henttonen H, Hantula J, Gubanyi HA (2004) Molecular and morphological evidence for multiple species within *Paranoplocephala omphalodes* (Cestoda, Anoplocephalidae) in *Microtus* voles (Arvicolinae). Zool Scr 33:277–290

Hedrick PW, Lee RN, Parker KM (2000a) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. Heredity 85:617–624

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

Higasa K, Kukita Y, Baba S, Hayashi K (2002) Software for machine-independent quantitative interpretation of SSCP in capillary array electrophoresis (QUISCA). Biotechniques 33:1342–1348

Hohenhaus MA, Outteridge PM (1995) The immunogenetics of resistance to *Trichostrongylus colubriformis* and *Haemonchus contortus* parasites in sheep. Br Vet J 151:119–140

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

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B 256:119–124

Hughes AL, Nei M (1990) Evolutionary relationships of class II MHC genes in mammals. Mol Biol Evol 7:491–514

Hughes AL, Yeager M (1998a) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32:415–435

Hughes AL, Yeager M (1998b) Natural selection and the evolutionary history of MHC loci. Front Biosci 3:509–516

Hughes AL, Hughes MK, Howell CY, Nei M (1994) Natural selection at the class II major histocompatibility complex loci of mammals. Philos Trans R Soc Lond B 346:359–366

Hurt P, Walter L, Sudbrak R, Klages S, Müller I, Shiina T, Inoko H, Lehrach H, Günther E, Reinhardt R, Himmelbauer H (2004) The genomic sequence and comparative analysis of the rat major histocompatibility complex. Genome Res 14:631–639

Jaarola M, Martinkova N, Gunduz I, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, Gonzalez-Esteban J, Lopez-Fuster MJ, Kandaurov AS, Kefelioglu H, Mathias MD, Villate I, Searle JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 33:647–663

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

Karlin S, Altschul SF (1990) Methods for assessing the statistical significance of molecular sequences by using general scoring schemes. Proc Natl Acad Sci U S A 87:2264–2268

Kennedy LJ, Ryvar R, Gaskell RM, Addie DD, Willoughby K, Carter SD, Thomson W, Ollier WER, Radford AD (2002) Sequence analysis of MHC *DRB* alleles in domestic cats from the United Kingdom. Immunogenetics 54:348–352

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

Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19:155–162

Klein J, O'Huigin C (1994) MHC polymorphism and parasites. Philos Trans R Soc Lond B 346:351–358

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

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

Latek RR, Suri A, Petzold SJ, Nelson CA, Kanagawa O, Unanue ER, Fremont DH (2000) Structural basis of peptide binding and presentation by the type I diabetes-associated MHC class II molecule of NOD mice. Immunity 12:699–710

Martin Y, Gerlach G, Schlötterer C, Meyer A (2000) Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. Mol Phylogenet Evol 16(1):37–47

Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. Immunogenetics 50:168–200

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

Meyer-Lucht Y, Sommer S (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Mol Ecol 14:2233–2243

Michaux J, Catzeflis F (2000) The bushlike radiation of muroid rodents is exemplified by the molecular phylogeny of the LCAT nuclear gene. Mol Phylogenet Evol 17:280–293

Musolf K, Meyer-Lucht Y, Sommer S (2004) Evolution of MHC-*DRB* class II polymorphism in the genus *Apodemus* and a comparison of *DRB* sequences within the family Muridae (Mammalia: Rodentia). Immunogenetics 56:420–426

Musser GG, Carleton MD (1993) Family Muridae. In: Wilson DE, Reeder DM (eds) Mammal species of the world: a taxonomic and geographic reference, 2nd edn. Smithsonian Institution, Washington, pp 501–755

Nei M, Gojobori T (1986) Simple method for estimating the numbers of synonymous and nonsynonymous 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

Ohta T (1999) Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunol Rev 167: 319–325

Otting N, de Groot NG, Doxiadis GGM, Bontrop RE (2002) Extensive Mhc-*DQB* variation in humans and non-human primate species. Immunogenetics 54:230–239

Ottova E, Simkova A, Martin JF, Gouy de Bellocq J, Gelnar M, Allienne JF, Morand S (2005) Evolution and trans-species polymorphism of MHC class IIβ genes in cyprinid fish. Fish Shellfish Immunol 18:199–222

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. Proc Natl Acad Sci U S A 95:3714–3719

Pfau RS, Van Den Bussche RA, McBee K, Lochmiller RL (1999) Allelic diversity at the Mhc-*DQA* locus in cotton rats (*Sigmodon hispidus*) and a comparison of *DQA* sequences within the family Muridae (Mammalia: Rodentia). Immunogenetics 49:886–893

Pfau RS, Van Den Bussche RA, McBee K (2001) Population genetics of the hispid cotton rat (*Sigmodon hispidus*): patterns of genetic diversity at the major histocompatibility complex. Mol Ecol 10:1939–1945

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol 19: 708–717

Quinnell RJ (2003) Genetics of susceptibility to human helminth infection. Int J Parasitol 33:1219–1231

Reusch TBH, Schaschl H, Wegner KM (2004) Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. Immunogenetics 56:427–437

Richman AD, Herrera LG, Nash D (2002) Characterization of *Peromyscus* MHC class II beta sequences by ligation-anchored RT-PCR and denaturing gradient gel electrophoresis. Eur J Immunogenet 29:213–217

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

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

Robinson J, Waller MJ, Stoehr P, Marsh SGE (2005) IPD—the immuno polymorphism database. Nucleic Acids Res 331: 523–526

Roitt IM (1997) Essential immunology, 9th edn. Blackwell Science, Oxford

Sauermann U, Nurnberg P, Bercovitch FB, Berard JD, Trefilov A, Widdig A, Kessler M, Schmidtke J, Krawczak M (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. Hum Genet 108:249–254

Sawyer SA (1999) GENECONV: a computer package for statistical detection of gene conversion. Available at http://www.math.wustl.edu/~sawyer/mbprogs/

Schad J, Ganzhorn JU, Sommer S (2005) Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. Evolution 59:439–450

Schaschl H, Suchentruk 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

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

Scott CA, Peterson PA, Teyton L, Wilson IA (1998) Crystal structures of two I-A^d–peptide complexes reveal that high affinity can be achieved without large anchor residues. Immunity 8:319–329

Seddon JM, Baverstock PR (1998) Eight rat RT1Ba sequences. Immunogenetics 48:161–162

Seddon JM, Baverstock PR (1999) Variation on islands: major histocompatibility complex (Mhc) polymorphism in populations of the Australian bush rat. Mol Ecol 8:2071–2079

Seddon JM, Baverstock PR (2000) Evolutionary lineages of *RT1.Ba* in the Australian *Rattus*. Mol Biol Evol 17:768–772

Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. Immunogenetics 54:490–500

Smulders MJM, Snoek LB, Booy G, Vosman B (2003) Complete loss of MHC genetic diversity in the common hamster (*Cricetus cricetus*) populations in The Netherlands. Consequences for conservation strategies. Cons Gen 4:441–451

Snibson KJ, Maddox JF, Fabb SA, Brandon MR (1998) Allelic variation of ovine MHC class II *DQA1* and *DQA2* genes. Anim Genet 29:356–362

Sommer S, Tichy H (1999) Major histocompatibility complex (MHC) class II polymorphism and paternity in the monogamous *Hypogeomys antimena*, the endangered, largest endemic Malagasy rodent. Mol Ecol 8:1259–1272

Sommer S, Schwab D, Ganzhorn JU (2002) MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. Behav Ecol Sociobiol 51:214–221

Steppan SJ, Adkins RM, Anderson J (2004) Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol 53:533–553

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

The MHC sequencing consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921–923

Vestberg M, Brunsberg U, Bergsteinsdottir K, Karlsson M, Gustafsson K, Wedekind D, Hedrich H, Holmdahl R (1998) Limited polymorphism in the first domain of the rat MHC class II *RT1-D* molecule. Immunogenetics 48:344–349

Waine GJ, Ross AG, Williams GM, Sleigh AC, McManus DP (1998) HLA class II antigens are associated with resistance or susceptibility to hepatosplenic disease in a Chinese population infected with *Schistosoma japonicum*. Int J Parasitol 28: 537–542

Wittzell H, Madsen T, Westerdahl H, Shine R, von Schantz T (1999) MHC variation in birds and reptiles. Genetica 104:301–309

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

Yang ZH, Swanson WJ, Vacquier VD (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol 17:1446–1455

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118