



Title	Genetic diversity of MHC class II DRB alleles in the continental and Japanese populations of the sable <i>Martes zibellina</i> (Mustelidae, Carnivora, Mammalia)
Author(s)	Nishita, Yoshinori; Abramov, Alexei V; Murakami, Takahiro; Masuda, Ryuichi
Citation	Mammal Research, 63(3), 369-378 <a href="https://doi.org/10.1007/s13364-018-0366-0">https://doi.org/10.1007/s13364-018-0366-0</a>
Issue Date	2018-07
Doc URL	<a href="http://hdl.handle.net/2115/74848">http://hdl.handle.net/2115/74848</a>
Rights	The original publication is available at <a href="http://www.springerlink.com">www.springerlink.com</a>
Type	article (author version)
File Information	USCAP_Sci Bio Nishita.pdf



[Instructions for use](#)

**Genetic Diversity of MHC Class II *DRB* Alleles in the Continental and Japanese Populations of the**

**Sable, *Martes zibellina* (Mustelidae, Carnivora, Mammalia)**

Yoshinori Nishita<sup>1</sup>, Alexei V. Abramov<sup>2</sup>, Takahiro Murakami<sup>3</sup>, Ryuichi Masuda<sup>1</sup>

<sup>1</sup>Department of Biological Sciences, Faculty of Science, Hokkaido University, North 10 West 8, Kita-ku,

Sapporo 060-0810, Japan

<sup>2</sup>Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, 199034 Saint Petersburg,

Russia

<sup>3</sup>Shiretoko Museum, 49-2 Hon-machi, Shari-cho, Hokkaido 099-4113, Japan

e-mail address and telephone number of the corresponding author, Yoshinori Nishita

e-mail: nishita@sci.hokudai.ac.jp

telephone: +81-11-806-2881

## Abstract

The sable (*Martes zibellina*) is a medium-sized mustelid inhabiting forest environments in Siberia, northern China, the Korean Peninsula, and Hokkaido Island, Japan. To further understand the molecular evolution of the major histocompatibility complex (MHC), we sequenced part of exon 2 in MHC class II DRB genes, including codons encoding the antigen binding site, from 33 individuals from continental Eurasia and Japan. We identified 16 MHC class II DRB alleles (*Mazi-DRBs*), some of which were geographically restricted and others broadly distributed, and eight putative pseudogenes. A single-breakpoint recombination analysis detected a recombination site in the middle of exon 2. A mixed effects model of evolution analysis identified five amino acid sites presumably under positive selection. These sites were all located in the region 3' to the recombination site, suggesting that positive selection and recombination could be committed to the diversity of the *M. zibellina* DRB gene. In a Bayesian phylogenetic tree, all *Mazi-DRBs* and the presumed pseudogenes grouped within a Mustelidae clade. The *Mazi-DRBs* showed trans-species polymorphism, with some alleles most closely related to alleles from other mustelid species. This result suggests that the sable DRBs have evolved under long-lasting balancing selection.

**Keywords**

Balancing selection · DRB gene · Major histocompatibility complex · Molecular evolution ·

Trans-species polymorphism · *Martes zibellina*

## Introduction

The sable (*Martes zibellina*) is a marten species that inhabits forest environments, in Russia, from the Ural Mountains throughout Siberia to Far East, eastern Kazakhstan northern Mongolia, China, the Korean Peninsula, and Hokkaido Island, Japan. Continental sables measure 35–56 cm in body length, with a 7.2–12 cm tail, and weigh 700–1800 g. Due to geographical variation in body size and fur color, which ranges from light to dark brown, various researchers have recognized anywhere from seven to thirty subspecies of sable (Monakhov 2011), but debate over the number of subspecies continues. Individuals in the insular subspecies *M. z. brachyura* inhabiting Hokkaido Island are smaller than in continental subspecies, measuring 40–50 cm in body length and weighing 1000–1300 g, but have a longer tail (10–15 cm); they are marked with black on their legs and feet (Miyoshi and Higashi 2005).

Previous phylogeographical studies of *M. zibellina* have used neutral markers such as mitochondrial DNA (mtDNA) and microsatellites to estimate the genetic diversity and distances within and among populations (Inoue et al. 2010; Nagai et al. 2012, 2014). Sequence data from neutral molecular markers can provide insights into genetic diversity, evolutionary potential, effective population size, and taxonomy (Ballentine and Greenberg, 2010; Kawakami et al. 2011; Richter-Boix et al. 2011). To better understand adaptation, information is necessary on the diversity of non-neutral genes directly

influencing fitness. Recent technological advances that allow the examination of adaptive genetic markers at hundreds or thousands of loci, combined with ongoing improvements in statistical tools and software for analyzing data from non-neutral markers, have led to studies that target regions and genes under the influence of selection (Kirk and Freeland 2011).

Because of high allelic diversity, the major histocompatibility complex (MHC) has attracted the attention of many evolutionary biologists. MHC gene families, found in all vertebrates, are subdivided into classes I, II, and III. The genes in classes I and II encode cell-surface glycoproteins that recognize peptides derived from intra- or extracellular pathogens, respectively, and present them to T-cells; these genes and the proteins they encode are thus crucial to the acquired immune system (Swain 1983; Toh et al. 2000; Abbas and Lichtman, 2009). MHC polymorphism protects a population against a new pathogen or a mutated one, because at least some individuals will be able to develop an adequate immune response to resist the pathogen. The variation in MHC molecules results from the inheritance of different MHC molecules, rather than by recombination, as is the case for the antigen receptors. A high level of MHC polymorphism has been attributed to positive selection in populations, with balancing selection maintaining multiple alleles at a particular locus within individuals (Doherty and Zinkernagel 1975; Hughes and Yeager 1998; Penn et al. 2002; Piertney and Oliver 2006). Some mammalian MHC allelic

lineages are tens of millions of years old, often predating the divergence of the species carrying the alleles (Figueroa et al. 1988). Consequently, phylogenetic reconstruction often reveals trans-species polymorphism, with some alleles from different species more closely related to one another than to other alleles within either species (Klein 1987; Klein and O’Huigin 1994; Edwards and Hedrick 1998; Klein et al. 1998).

Within a  $\beta$ -chain of MHC class II DR proteins, the residues that comprise the antigen-binding site (ABS) in the  $\beta$ 1-domain are encoded by *DRB* exon 2; the majority of the polymorphic sites contributing to allelic diversity in *DRB* reside in exon 2, especially in the ABS codons (Hughes & Yeager 1998). In the present study, we sequenced a 242-base-pair (bp) region within *DRB* exon 2 that contributes to the ABS and analysed this region for evidence of selection at the codon level. We also reconstructed the phylogenetic relationships among *DRB* exon 2 sequences for evidence of trans-species polymorphism, which is indirect evidence for balancing selection. Based on our results, we discuss the level of divergence in the MHC class II DRB gene between Eurasian and Japanese (Hokkaido) populations of *M. zibellina* at the species level, relative to that in other mustelid species.

## Material and methods

### Samples and molecular methods

Muscle tissue samples were collected from 33 *M. zibellina* individuals kept in the collections of the Obihiro University of Agriculture and Veterinary Medicine, Shiretoko Museum and the Zoological Institute of the Russian Academy of Sciences (Fig. 1; Table 1) and preserved in 99.5% ethanol. Total DNA was extracted from the samples by using the DNeasy Blood and Tissue Kit (Qiagen, Venlo, the Netherlands) according to the manufacturer's protocol, and was stored at 4°C until analysis.

Part of MHC class II *DRB* exon 2 was amplified from *M. zibellina* by polymerase chain reaction (PCR) using a consensus forward primer designed from *Meles meles* (Meme-DRBex2F; 5'-CGTCCCCACAGGACATTTC-3'; Sin et al. 2012) and a consensus reverse primer designed from *Mustela lutreola* (*M. lutreola* reverse primer; 5'-CTCGCCGCTGCACCGTGAAG-3'; Becker et al. 2009). PCR amplifications were performed in 25 µl reaction volumes, each containing 1x Buffer (Mg<sup>2+</sup> plus), 0.2 nmol of dNTP, 25–50 ng of the DNA extract as template, 7.5 pmol of each phosphorylated primers, and 0.625 U of PrimeSTAR<sup>®</sup> GXL DNA polymerase (Takara Bio, Kusatsu, Japan). Reactions were run in a Takara Dice<sup>®</sup> Touch thermal cycler PCR (Takara Bio). PCR conditions were 94°C for 2 min; 30 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 30 s; and 68°C for 5 min. To check molecular sizes, PCR products were separated on a 3% agarose gel and visualised with ethidium-bromide fluorescence.

PCR products amplified from *M. zibellina* MHC class II *DRB* exon 2 (281 bp, including primer sequences) were purified with the QIAquick PCR Purification Kit (Qiagen), cloned into pBluescript II SK+ (Agilent Technologies, Santa Clara, CA, USA), and transformed into *Escherichia coli* JM109 competent cells. Positive clones, containing plasmids with target amplicons, were identified by blue/white selection and direct-colony PCR. Plasmids were isolated from cultures by using NucleoSpin<sup>®</sup> Plasmid EasyPure columns (Macherey-Nagel GmbH, Düren, Germany). Positive clones were randomly selected and sequenced in both directions with M13 forward and reverse primers (Messing 1983) by using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific K. K., Yokohama, Japan) and an ABI3730 automated DNA sequencer (Applied Biosystems, Carlsbad, CA, USA). Between 32 and 48 clones were sequenced for each individual.

### **Data analysis**

After manually trimming of the vector and primer sequences, the nucleotide sequences obtained were aligned using GENETYX<sup>®</sup>-MAC Ver.19 (GENETYX, Tokyo, Japan). Sequences were confirmed as MHC class II *DRB* exon 2 or previously reported pseudogenes by means of BLAST-N and -X searches (Altschul et al. 1990) of the NCBI GenBank database. Insert sequences were identified as bona-fide MHC

variants if the forward and reverse sequences were consistent, and identical sequences were obtained from at least two individuals or from two independent PCR reactions from the same individual. Single, unique sequences were excluded, as they may have been of chimeric origin. Verified sequences were named according to the nomenclature rules for the dog MHC (DLA; dog leukocyte antigen) (Kennedy et al. 1999), with the species prefix followed by *DRB* (the abbreviation for the gene encoding the MHC class II DR  $\beta$ -chain) and an identification number, as recommended by Klein et al. (1990). Our data set for calculating diversity statistics consisted of one copy of each allele from each individual, regardless of the frequencies among clone sequences. To measure sequence polymorphism,  $\pi$  (nucleotide diversity; average number of pairwise differences),  $\theta$  ( $4N\mu$  for an autosomal gene in a diploid organism), and Tajima's  $D$  were calculated from the allele dataset by using DnaSP (Librado and Rozas 2009). Tajima's  $D$  is a statistical test to distinguish between a DNA sequence evolving randomly and one evolving under a non-random process such as directional or balancing selection (Tajima 1989).

The ABS codons were examined for signs of positive selection. The ratio of non-synonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitutions ( $\omega=d_N/d_S$ ) is an important indicator of selective pressure at the amino acid level (Nei and Kumar 2000; Yang and Nielsen 2002), with  $\omega=1$  indicating neutral mutations,  $\omega < 1$  purifying selection, and  $\omega > 1$  diversifying positive selection. To estimate  $\omega$ -values, the average

frequencies of synonymous and non-synonymous substitutions per site were calculated by using the Nei-Gojobori method (Nei and Gojobori 1986). The 22 codons in *M. zibellina DRB* exon 2, those are homologous to the ABS codons in the human leukocyte antigen (HLA) complex (the human equivalent of the MHC) (Toh et al. 2000) determined from the three-dimensional crystallographic structure of HLA-DR1 (Brown et al. 1993), were considered to encode ABSs of *M. zibellina DR*  $\beta$ -chain (Fig. 2). To test for selection,  $\omega$ -values were then calculated for ABS positions, non-ABS positions, and all positions. The analysis was performed with MEGA 6.0 (Tamura et al. 2011). In order to obtain detailed data on sites under positive selection and recombination break points, additional analyses were performed using the mixed effects model of evolution (MEME), Genetic Algorithm for Recombination Detection (GARD) and single breakpoint recombination (SBP) methods provided by DATAMONKEY, a web-server for the HyPhy Package (Murrell et al. 2012).

A Bayesian phylogenetic analysis was performed to reconstruct the evolutionary relationships among alleles detected in *M. zibellina* and those in other species in Carnivora, including other Mustelidae, *Canis*, and *Felis*, obtained from GenBank (Fig. 3: accession numbers in parentheses), and to investigate whether *DRB* alleles in Mustelidae exhibit trans-species polymorphism. Sequences were aligned by using GENETYX<sup>®</sup>-MAC, with Gblocks (Talavera and Castresana 2007) selected to give the greatest number of

informative nucleotide positions under the less stringent option. Kakusan4 (Tanabe, 2011) under the Akaike information criterion (AIC) selected HKY85\_Gamma (1st and 3rd partitions) and F81\_Gamma (2nd partition) as the optimal substitution models. The Bayesian phylogeny was inferred by using MrBayes v3.2.6 (Ronquist and Huelsenbeck 2003) with two Markov chain Monte Carlo (MCMC) simulations each run for  $5 \times 10^8$  generations and trees sampled every 1000 generations. The convergence of parameter values sampled from the chains was checked with Tracer v1.6.0 (Rambaut et al. 2014), and first 10% of generations were discarded as burn-in.

## **Results**

### **Diversity of MHC *DRB* alleles in *Martes zibellina***

We obtained 242-bp nucleotide sequences of MHC class II *DRB* exon 2, corresponding to a region encoding 80 amino acids, or 84% of the  $\beta$ 1-domain (total length 285 bp, encoding 95 amino acids), as well as some other sequences having pseudogene features. From the 33 individuals of *M. zibellina*, we sequenced 1051 clones (average 31.8 clones per individual). Among the sequences, we identified 16 distinct nucleotide sequences, each containing one open reading frame, as MHC class II *DRB* alleles, which we named *Mazi-DRB\*01-16* (DDBJ/EMBL/GenBank accession numbers LC309250–LC309265).

By contrast, eight sequences, Mazi-DRB\*PS1–8 (DDBJ/EMBL/GenBank accession numbers LC309266–LC309273), appeared to be pseudogenes; these had stop codons or indels in the protein coding region and high similarity to previously reported pseudogene sequences. Among the alleles, BLAST search identified six *Mazi-DRBs* and two pseudogenes that were identical to sequences from *Mustela nivalis* (*Muni-DRBs*): *Mazi-DRB*\*01, \*02, \*03, \*04, \*05, \*06, \*PS2, and \*PS4 were identical to *Muni-DRB*\*04, \*16, \*02, \*06, \*03, \*01, \*PS03, and \*PS02, respectively. Except for the presumed pseudogenes, all the *Mazi-DRBs* translated to distinct amino acid sequences (Fig. 2).

### **Distribution of MHC DRB alleles**

Alleles *Mazi-DRB*\*01–07 were found in *M. zibellina* individuals from two or more locations both on the continent and in Japan. All individuals shared *Mazi-DRB*\*01 regardless of sampling location (Table 1), and *Mazi-DRB*\*02 was the second most frequently found in individuals from geographically wide areas. By contrast, *Mazi-DRB*\*08–13 were detected only in continental individuals, and *Mazi-DRB*\*14–16 only in Japanese individuals. Among the eight presumed pseudogenes, various mutations led to frame shifts or premature stop codons: *Mazi-DRB*\*PS1, -3, -5, -7, and -8 had a one-bp deletion, *Mazi-DRB*\*PS2 and -4 contained an insertion, and *Mazi-DRB*\*PS6 showed a truncated region. Pseudogenes *Mazi-DRB*\*PS1–3

were detected in individuals both on the continent and in Japan, whereas the others were found only in continental individuals. None of the presumed pseudogenes was restricted to Japan. Aside from the presumed pseudogenes, the number of alleles detected per individual ranged from two (individuals Obihiro A, Shiretoko C and F) to eight (Shiretoko G). This indicates that there are one to four MHC *DRB* loci per haploid genome in *M. zibellina*.

#### **Analysis of selection on *DRB* exon 2**

To infer the direction and magnitude of natural selection acting on *M. zibellina DRBs*, the ratio ( $\omega$ ) of the non-synonymous substitution rate ( $d_N$ ) to the synonymous substitution rate ( $d_S$ ) was calculated. For the predicted ABS codons (Fig. 2), non-ABS codons, and all codons (ABS and non-ABS), the rate of synonymous substitutions exceeded that of non-synonymous substitutions ( $\omega=0.783, 0.396, \text{ and } 0.592$ , respectively). Although the  $\omega$ -value was higher for ABS codons than for non-ABS codons or all positions, the value less than 1 implies that purifying or stabilizing selection acts on the *M. zibellina DRB* locus. On the other hand, the MEME analysis indicated positive selection at the level of individual codons in exon 2. Five codons in the 242-bp region of *Mazi-DRBs* showed significant ( $P < 0.05$ ) evidence for positive selection (asterisks in Fig. 2). Most of these were congruent with ABS codons, although one of the amino

acids represented were located next to ABS position. Finally, the value of 1.17 for Tajima's  $D$  ( $\pi = 0.088$ ,  $\theta$  from  $S = 0.064$ ; not significant,  $P > 0.10$ ) was greater than zero, indicating that balancing selection has acted on the *DRB* loci.

Eighty-four potential recombination breakpoints in the 242-bp region were found based on beta-gamma rate variation or the general discrete rate variation model (2–6 rate classes) in the SBP analysis (Kosakovsky Pond et al. 2006), and the same result was obtained using GARD. Among them, one recombination breakpoint was detected at nucleotide position 119 (amino acid position 47 in the  $\beta 1$  domain; triangle in Fig. 2) with a model-averaged support value of 100% [AIC/ small sample AIC (cAIC)/BIC].

### **Phylogenetic relationships and trans-species polymorphism among MHC class II *DRB* alleles**

A Bayesian phylogenetic tree (Fig. 3) shows sequences from *M. zibellina* grouped within a clade comprising Mustelidae. All mustelid sequences form a large, unresolved polytomy that includes nine subclades (A–H and B'), with variable nodal support (posterior probability values of 0.51–1); six of the nine subclades contained *Mazi-DRB* sequences. All but one subclade contains *DRB* alleles from two or more mustelid species. Only one subclade (H) comprises exclusively *M. zibellina DRB* alleles,

*Mazi-DRB\*10* and *11*, with only one nucleotide substitution between them, and both of which occurred in individuals from Surgut, Russia. Other *Mazi-DRBs* are distributed among subclades A, B', C, D, and G. All eight presumed pseudogenes (*Mazi-DRB-PS\*1-8*) we detected are in subclade B', along with three presumed pseudogenes from *Mustela nivalis*. In all three families of Carnivora analysed in the present study (Mustelidae, Felidae, and Canidae), alleles from some species are more closely related to those from another species than to those from the same species, indicating trans-species polymorphism.

## **Discussion**

### **MHC diversity in *M. zibellina***

In the present study, 16 *DRB* alleles and eight pseudogene sequences were identified in *M. zibellina*. This number of alleles is similar to that in other mustelid species previously analysed (Nishita et al. 2015, 2017).

For example, 32 sequences (27 *DRB* alleles and five pseudogene sequences) detected from 35 *Mustela nivalis* individuals, 25 sequences (24 *DRB* alleles and one pseudogene sequence) from 31 Japanese weasels *Mustela itatsi*, and 17 sequences (16 *DRB* alleles and one pseudogene sequence) from 21

Siberian weasels *Mustela sibirica*. Other studies have detected fewer alleles in mustelids: four in *Meles*

*meles* (Sin et al. 2012; Abduriyim et al. 2017), nine in *Mustela lutreola* (Becker et al. 2009), and nine in the wolverine *Gulo gulo* (including *DRB1* and *DRB2*) (Oomen et al. 2013).

Among 16 *Mazi-DRB\** alleles detected, seven occurred in both continental and Japanese populations, six were restricted to continental populations, and three were restricted to Japanese populations. *Mazi-DRB\*01* was detected in all *M. zibellina* individuals sampled and could have originated from a separate, probably monomorphic or limited-polymorphic MHC locus such as HLA-E, F, or G of non-classical human MHC class I (class Ib) (Kumánovics et al. 2003). The varied distributions of alleles we observed could be critical for immune defence: widespread alleles could defend against widespread pathogens, whereas divergent, locally restricted alleles (reflecting variation mainly in the ABS) could be maintained by pathogen-driven balancing selection and defend against native pathogens in particular areas (Hughes and Yeager 1998).

Relative to the number of the *DRB* alleles, we detected a high number of pseudogenes in *M. zibellina*. This result seems to be concordant with  $\omega$ -values less than 1, indicating that *Mazi-DRBs* are under purifying selection. Five pseudogene sequences (*Mazi-DRB\*PS4–8*) were detected only in continental Eurasian populations, three (*Mazi-DRB\*PS1–3*) were detected in both continental and Japanese populations, but none were restricted to Japanese subpopulation, *Martes zibellina brachyuran*.

Among *Mazi-DRBs* and the pseudogenes we detected, six alleles and two pseudogenes were identical in nucleotide sequence to alleles or pseudogenes in *Mustela nivalis*, although *Martes zibellina* and *Mustela nivalis* are classified in different subfamilies in Mustelidae.

The numbers of alleles detected ranged from two and eight (one and four loci) per individual, suggesting existence of gene loci with the same nucleotide sequence and/or that the number of MHC *DRB* loci ranges from one to four per haplotype genome in *M. zibellina*. Similar numbers of loci have been estimated in other mustelid species: one to three in *Mustela itatsi*, *Mustela sibirica*, *Mustela nivalis*, and *Meles meles*; two or three in *Meles leucurus* and *Meles anakuma*; and four in *Meles canescens* (Nishita et al., 2015, 2017; Abduriyim et al., 2017). Other mammalian species, including humans (Traheme, 2008), also show variation in the number (CNV) of *DRB* loci among individuals (Doxiadis et al., 2010) as well as between species (Kelley et al., 2005).

### **Selection on *DRB* exon 2**

Evolution acts on protein-coding genes through selection on protein variants. Mutations leading to synonymous codons are largely invisible to natural selection, whereas non-synonymous mutations may be under strong selective pressure. The  $\omega$ -values ( $d_N/d_S$ ) we obtained for ABS, non-ABS, and all codons

were consistently less than 1, suggesting that *Mazi-DRBs* evolved under purifying or stabilising selection.

However, a mixture of advantageous and disadvantageous mutations could also result in values less than

1 (Yang and Bielawski 2000). Partly inconsistent with the  $\omega$ -values we obtained, the MEME analysis

showed some amino acids encoded by the 242-bp region to be under positive selection. Overall, our data

strongly suggest that positive selection acts on *M. zibellina DRBs*. The SBP analysis detected a

recombination breakpoint at nucleotide position 119 from the 5' end of the 242-bp region in *Mazi-DRB*

exon 2. Because the sites predicted to be under positive selection were all 3' to the recombination site, it

is possible that *M. zibellina DRB* alleles contain both paralogous *Mazi-DRB* sequences and other,

non-polymorphic genes through inter-locus recombination. This suggests that recombination events could

have occurred between different alleles, helping maintain genetic diversity (Martinsohn et al. 1999).

Reche and Reinherz (2003) reported a correlation between ABS residues and positively selected residues

in human class II MHC molecules. Most of the positively selected residues in *M. zibellina* corresponded

with those in the human class II MHC; however, some residues were different between the two species.

This suggests that we could have misidentified some ABS positions in *M. zibellina DRBs*, because we

estimated them by homology with those in a human MHC class II molecule, HLA-DR1, and thus the

$\omega$ -value could incorporate a certain amount of error. As our data were only for part of exon 2, a future analysis of the entire exon would be more informative.

### **Phylogenetic relationships among *DRB* alleles**

Family Mustelidae, with eight subfamilies, is the largest, most diverse, and most widely distributed family in order Carnivora. However, the internal classification has been controversial, with alternative proposals existing for between two and eight subfamilies. Different phylogenetic analyses, for example, based on the paternally inherited SRY and ZFY gene, the mitochondrial cytochrome *b* gene, autosomal and X-linked genes, have indicated different relationships among mustelid subfamilies (Koepli et al. 2008; Sato et al. 2012; Mizumachi et al. 2017).

The topology of the phylogenetic tree reconstructed in the present study is similar to that reported in Nishita et al. (2017). The phylogenetic tree in a previous study (Nishita et al. 2017) included subclade C containing only *Gulo gulo* *DRB* alleles; the corresponding clade in our tree includes not only all the *G. gulo* *DRBs* but also *Meme-DRB\*04* and *Mazi-DRBs* (\*07, \*08, \*13 and \*16). This grouping of *DRB* alleles from *M. zibellina* and *G. gulo* makes sense, because these species are also shown to be relatively closely related according to the analyses based on neutral genetic markers and are together

classified into the same subfamily, Guloninae (Sato et al. 2012). Clade B in Nishita et al. (2017) was, in our tree, divided into subclades B (containing mostly *Mustela nivalis* DRB alleles) and B' (containing the presumed *M. zibellina* pseudogenes as well as *Mazi-DRB\*09* and *\*15*); therefore, the pseudogenes in *M. zibellina* DRB might all have derived from *Mazi-DRB\*09* or *\*15*. Subclade A includes DRB alleles from most species, indicating that these alleles encode a DR  $\beta$ -chain having an important immunological role across mustelids.

Our tree shows several instances of trans-species polymorphism among *Mazi-DRB* alleles. In these cases, allelic lineages present ancestrally species have been retained by balancing selection even after a common ancestor split into two or more descendant species (Klein 1987; Klein and O'Huigin 1994; Edwards and Hedrick 1998). This result is congruent with the Tajima's *D* value for *Mazi-DRBs* (1.17; greater than zero), though not statistically significant, showing that natural selection to maintain genetic variation, that is, balancing selection, appears to have acted on the *M. zibellina* DRB genes.

Further studies on MHC genes from other mustelid species could illustrate more precisely the relationships between species divergences and the diversification of MHC genes among species.

## Acknowledgements

We thank Dr Hisashi Yanagawa (Obihiro University of Agriculture and Veterinary Medicine), Dr Yasushi Masuda (Shiretoko Museum), and Ms. Tatiana Bulyonkova (Novosibirsk) for providing specimens, and Dr Matthew H. Dick for commenting on and editing the manuscript. This study was supported in part by a Joint Research Project Grant from the Japan Society for the Promotion of Science (JSPS) and the Russian Foundation for the Basic Research (RFBR 16-54-50004), and a Joint Research Program Grant from the Japan Arctic Research Network Centre.

## References

- Abbas AK, Lichtman AHH (2009) *Basic Immunology: Functions and Disorders of the Immune System*, 3rd edn. Saunders, Philadelphia
- Abduriyim S, Nishita Y, Kosintsev PA, Raichev E, Väinölä R, Kryukov AP, Abramov AV, Kaneko Y, Masuda R (2017) Diversity and evolution of MHC class II DRB gene in the Eurasian badger genus *Meles* (Mammalia: Mustelidae). *Biol J Linn* 122:258–273. doi: 10.1093/biolinnean/blx077
- Altschul SF, Gish W, Miller W, Myers EW, Lipman D.J (1990) Basic local alignment search tool. *J Mol Biol* 215:403–4109. doi: 10.1016/S0022-2836(05)80360-2
- Ballentine B, Greenberg R (2010) Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. *PLoS One* 5:e10229. doi: 10.1371/journal.pone.0010229
- Becker L, Nieberg C, Jahreis K, Peters E (2009) MHC class II variation in the endangered European mink *Mustela lutreola* (L. 1761) – consequence for species conservation. *Immunogenetics* 61:281–288. doi: 10.1007/s00251-009-0362-2
- 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*

364:33–39. doi: 10.1038/364033a0

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, de Groot N, de Groot NG, Rotmans G, de Vos-Rouweler AJ, Bontrop RE (2010)

Extensive DRB region diversity in cynomolgus macaques: recombination as a driving force.

*Immunogenetics*. 62:137–147. doi: 10.1007/s00251-010-0422-7

Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual

selection. *Trends Ecol Evol* 13:305–311. doi: 10.1016/S0169-5347(98)01416-5

Figueroa F, Gunther E, Klein J (1988) MHC polymorphism predating speciation. *Nature* 335:265–267.

doi: 10.1038/335265a0

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates.

*Annu Rev Genet* 32:415–435. doi: 10.1146/annurev.genet.32.1.415

Inoue T, Murakami T, Abramov AV, Masuda R (2010) Mitochondrial DNA control region variations in

the sable *Martes zibellina* of Hokkaido Island and the Eurasian continent, compared with the

Japanese marten *M. melampus*. *Mammal Study* 35:145–155. doi: 10.3106/041.035.0301

Kawakami T, Morgan TJ, Nippert BJ, Ocheltree TW, Keith R, Dhakal P, Ungerer MC (2011) Natural

selection drives clinal life history patterns in the perennial sunflower species, *Helianthus*

*maximiliani*. *Mol Ecol* 20:2318–2328. doi: 10.1111/j.1365-294X.2011.05105.x

Kennedy LJ, Altet L, Amgles JM, Barnes A, Carter SD, Francino O, Gerlach J A, Happ GM, Ollier WER,

Polvi A, Thomson W, Wagner JL (1999) Nomenclature for factors of the dog major

histocompatibility system (DLA), 1998. First report of the ISAG DLA Nomenclature Committee.

*Tissue Antigens* 54:312–321. doi: 10.1034/j.1399-0039.1999.540319.x

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes.

*Immunogenetics*. 56:683–695. doi: 10.1007/s00251-004-0717-7

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutral markers in

molecular ecology. *Int J Mol Sci* 12:3966–3988. doi: 10.3390/ijms12063966

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis.

*Hum Immunol* 19:155–162

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. doi: 10.1007/BF00204890

Klein J, O’Huigin C (1994) MHC polymorphism and parasites. *Philos T R Soc B* 346:351–358. doi:

10.1098/rstb.1994.0152

Klein J, Sato A, Nagl S, O’huigin C (1998) Molecular transspecies polymorphism. *Annu Rev Ecol Evol S*

29:1–21. doi: 10.1146/annurev.ecolsys.29.1.1

Koepfli KP, Deere KA, Slater GJ, Begg C, Begg, K, Grassman L, Lucherini M, Veron G, Wayne RK

(2008) Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic

history of a mammalian adaptive radiation. *BMC Biol* 6:1–22. doi: 10.1186/1741-7007-6-10

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic

detection of recombination using a genetic algorithm. *Mol Biol Evol* 23:1891–1901. doi:

doi.org/10.1093/molbev/msl051

Kumánovics A, Takada T, Lindahl KF (2003) Genomic organization of the mammalian MHC. *Annu Rev*

*Immunol* 21:629–657. doi: 10.1146/annurev.immunol.21.090501.080116

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.

*Bioinformatics* 25:1451–1452. doi: 10.1093/bioinformatics/btp187

Martinson JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC

evolution: a review. *Immunogenetics* 50:168–200. doi: doi.org/10.1007/s002510050

Messing J (1983) New M13 vectors for cloning. *Methods Enzymol* 101:20–78

Miyoshi K, Higashi S (2005) Home range and habitat use by the sable *Martes zibellina brachyura* in a Japanese cool-temperate mixed forest. *Ecol Res* 20:95–101. doi: 10.1007/s11284-004-0012-y

Mizumachi K, Nishita Y, Spassov N, Raichev EG, Peeva S, Kaneko Y, Masuda R. (2017) Molecular phylogenetic status of the Bulgarian marbled polecat (*Vormela peregusna*, Mustelidae, Carnivora), revealed by Y chromosomal genes and mitochondrial DNA sequences. *Biochem Syst Ecol.* 70:99–107. doi: 10.1016/j.bse.2017.01.011

Monakhov VG (2011) *Martes zibellina* (Carnivora: Mustelidae). *Mamm Spp* 43:75–86. doi: 10.1644/876.1

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764. doi: 10.1371/journal.pgen.1002764

Nagai T, Murakami T, Masuda R (2012) Genetic variation and population structure of the sable *Martes zibellina* on eastern Hokkaido, Japan, revealed by microsatellite analysis. *Mammal Study* 37:323–330. doi: 10.3106/041.037.0406

- Nagai T, Murakami T, Masuda R (2014) Effectiveness of noninvasive DNA analysis to reveal isolated-forest use by the sable *Martes zibellina* on eastern Hokkaido, Japan. *Mammal Study* 39:99–104. doi: 10.3106/041.039.0205
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York
- Nishita Y, Abramov AV, Kosintsev PA, Lin LK, Watanabe S, Yamazaki K, Kaneko Y, Masuda R (2015) Genetic variation of the MHC class II DRB genes in the Japanese weasel, *Mustela itatsi*, endemic to Japan, compared with the Siberian weasel, *Mustela sibirica*. *Tissue Antigens* 86:431–442. doi: 10.1111/tan.12700
- Nishita Y, Kosintsev PA, Haukisalmi V, Väinölä R, Raichev E, Murakami T, Abramov AV, Kaneko Y, Masuda R (2017) Diversity of MHC class II *DRB* alleles in the Eurasian population of the least weasel, *Mustela nivalis* (Mustelidae, Mammalia). *Biol J Linn Soc* 121:28–37. doi: 10.1093/biolinnean/blw028
- Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep

coverage. *Mol Ecol Resour* 13:103–116. doi: 10.1111/1755-0998.12027

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against

multiple-strain infections. *Proc Natl Acad Sci USA* 99:11260–11264. doi:10.1073/pnas.162006499

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex.

*Heredity* 96:7–21. doi: 10.1038/sj.hdy.6800724

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) *Tracer v1.6*. Available at:

<http://beast.bio.ed.ac.uk/Tracer> (accessed 11 December 2013)

Reche PA, Reinherz EL (2003) Sequence variability analysis of human class I and class II MHC

molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol*

15:623–641. doi: 10.1016/S0022-2836(03)00750-2

Richter-Boix A, Quintela M, Segelbacher G, Laurila A (2011) Genetic analysis of differentiation among

breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Mol Ecol*

20:1582–1600. doi: 10.1111/j.1365-294X.2011.05025.x

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models.

*Bioinformatics* 19:1572–1574. doi: 10.1093/bioinformatics/btg180

Sato JJ, Wolsan M, Prevosti FJ, D'Elía G, Begg C, Begg K, Hosoda T, Campbell KL, Suzuki H (2012)

Evolutionary and biogeographic history of weasel-like carnivorans (Musteloidea). *Mol Phylogenet*

*Evol* 63:745–757. doi: doi.org/10.1016/j.ympev.2012.02.025

Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T (2012) MHC class II genes in the European

badger (*Meles meles*): characterization, patterns of variation, and transcription analysis.

*Immunogenetics* 64:313–327. doi: 10.1007/s00251-011-0578-9

Swain S (1983) T cell subsets and the recognition of MHC class. *Immunol Rev* 74:129–142. doi:

10.1111/j.1600-065X.1983.tb01087.x

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.

*Genetics* 123:585–595

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously

aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577. doi:

10.1080/10635150701472164

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary

genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony

methods. *Mol Biol Evol* 28:2731–2739. doi: 10.1093/molbev/msr121

Tanabe AS (2011) Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional

and separate models for combined molecular phylogenetic analyses of multilocus sequence data.

Mol Ecol Resour 11:914–921. doi: 10.1111/j.1755-0998.2011.03021.x

Toh H, Savoie CJ, Kamikawaji N, Muta S, Sasazuki T, Kuhara S (2000) Changes at the floor of the

peptide-binding groove induce a strong preference for proline at position 3 of the bound peptide:

molecular dynamics simulations of HLA-A\*0217. Biopolymers 54:318–327. doi:

10.1002/1097-0282(20001015)54:5<318::AID-BIP30>3.0.CO;2-T

Traheme JA (2008) Human MHC architecture and evolution: implications for disease association studies.

Int J Immunogenet 35:179–192. doi: 10.1111/j.1744-313X.2008.00765.x

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol

15:496–503. doi: 10.1016/S0169-5347(00)01994-7

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual

sites along specific lineages. Mol Biol Evol 19:908–917. doi:

10.1093/oxfordjournals.molbev.a004148

### Figure legends

**Fig. 1** Map of Eurasia showing the sampling localities for 33 specimens of *Martes zibellina* analysed in the present study. The sample size at each locality is indicated in parenthesis.

**Fig. 2** Amino acid sequences deduced from the nucleotide sequences of part of *Martes zibellina* MHC class II *DRB* exon 2. Identity with the amino acid sequence encoded by *Mazi-DRB\*01* is indicated by dots. The numbers above the top sequence indicate codon positions in the  $\beta$ 1-domain. Grey shading shows residues in the antigen-binding site (ABS) predicted from data on the human gene HLA-DR1 (Brown et al. 1993). Asterisks indicate codons corresponding to sites under positive selection deduced from the mixed effects model of evolution (MEME) analysis. The triangle shows a recombination breakpoint determined with the single breakpoint recombination (SBP) method (see the text).

**Fig. 3** Bayesian phylogenetic reconstruction for MHC class II *DRB* alleles from *Martes zibellina* and other species in Mustelidae, Canidae and Felidae. Numerals near nodes are posterior probability values.

Sequences from *M. zibellina* obtained in the present study are in bold font. **The Canidae and Feridae**

clades are included as outgroups. GenBank accession numbers of previously published nucleotide sequences are in parentheses. Allele names that include 'PS' preceding the final numeral indicate putative pseudogenes. Subclades in the Mustelidae clade are indicated by capital letters (A to H and B') to the right of the tree. The scale bar at the bottom shows branch length in substitutions per site. Abbreviations for species names are as follows: Cafa, *Canis familiaris*; Cala, *Canis latrans*; Calu, *Canis lupus*; Enlu, *Enhydra lutris*; Febi, *Felis bieti*; Feca, *Felis catus*; Fesi, *Felis silvestris*; Gugu, *Gulo gulo*; Meme, *Meles meles*; Mui, *Mustela itatsi*; Mulu, *Mustela lutreola*; Muni, *Mustela nivalis*; Musi, *Mustela sibirica*; Nevi, *Neovison vison*; Tata, *Taxidea taxus*.

Table 1. Distribution of MHC class II *DRB* alleles and pseudogenes (PS) detected in 33 *Martes zibellina* individuals.

Sample Name	Sample ID (Collection No.)	<i>Mazi-DRB</i>																Number of alleles								
		*01	*02	*03	*04	*05	*06	*07	*08	*09	*10	*11	*12	*13	*14	*15	*16		*PS1	*PS2	*PS3	*PS4	*PS5	*PS6	*PS7	*PS8
Altai	(ZIN 97254)	+				+																				5
Barguzin A	(ZIN 76285)	+	+	+	+	+																				5
Barguzin B	(ZIN 76286)	+	+	+																	+					3 (4)
Barguzin C	(ZIN 76287)	+	+	+	+																+					4 (5)
Barguzin D	(ZIN 97253)	+	+			+	+														+		+			4 (6)
Dzhida	SPB67 (ZIN 35116)	+		+	+							+										+		+		4 (6)
Kamchatka	(ZIN 97079)	+	+		+		+													+						4 (5)
Sayan	(ZIN 97078)	+	+	+		+	+																+			5 (6)
Surgut A	SPB29	+	+							+	+									+	+					4 (6)
Surgut B	SPB30 (ZIN 37970)	+		+	+	+					+	+									+		+			6 (8)
Surgut C	SPB31 (ZIN 37971)	+	+		+					+	+										+					5 (6)
Surgut D	SPB32 (ZIN 37972)	+	+	+		+							+								+		+			5 (7)
Surgut E	SPB33 (ZIN 37973)	+	+	+						+																4
Surgut F	SPB34 (ZIN 37974)	+		+					+	+	+												+			5 (6)
Surgut G	SPB35 (ZIN 37975)	+	+								+	+		+								+		+		5 (7)
Kushiro	MZ-49	+	+	+										+												4
Obihiro A	MZI-CH2	+	+																							2
Obihiro B	MZI-CH4	+					+								+											3
Obihiro C	MZI-CH5	+	+		+		+																			4
Obihiro D	MZ-12	+			+			+																		3
Obihiro E	MZ-13	+	+	+																					+	4
Obihiro F	MZ-23	+	+			+																				3
Shiretoko A	MZ-2	+	+		+		+														+					5 (6)
Shiretoko B	MZ-4	+			+																	+	+			3 (4)
Shiretoko C	MZ-5	+	+																		+					2 (3)
Shiretoko D	MZ-8	+	+																							2
Shiretoko E	MZ-10	+	+	+		+																				4
Shiretoko F	MZ-29	+					+																			2
Shiretoko G	MZ-34	+	+	+	+	+	+							+		+	+									8 (9)
Shiretoko H	MZ-36	+	+												+		+	+	+							3 (6)
Shiretoko I	MZ-41	+	+			+											+									3 (4)
Shiretoko J	MZ-58	+	+														+									2 (3)
Shiretoko K	MZI-S1	+	+	+											+											4

Values in parentheses under 'Number of alleles' include the number of pseudogene sequences.

Asterisks are shown as a part of allelenames such as *Mazi-DRB*"\*01.

	10	20	30	40	50	60	70	80	
*01	LHL	TTSECYFTNG	TERVRFLE	FYNGEEFVRF	DSDVGEYRPV	TELGRRDAEY	WNSQKDIMED	ERATVDTYCR	HNYGVFE
*02	...	.....	.....Q..	.....D...	.....F.....	.....	.....	.....A.....	.....
*03	.Y.	.....	.....L..H	.....	.....	.....PS.QS	.....FL..	A..P.....	.....E.G.
*04	.L.	AKFQ.....	.....QL.IKA	V.....	.....F.....	.....PS.QS	.....FL..	A..P.....	.....E.G.
*05	.Y.	.....	.....L..H	.....	.....F.....	.....PS.Q.	.....FL..	A..P.....	.....E.G.
*06	.Y.	R.....	.....L..H	.....	.....	.....PS.QS	.....FL..	A..P.....	.....E.G.
*07	MEQ	FKA...H...	.....Y.V..	I..R..YA..	.N...F.....	.....PA.Q.	L.....L.Q	M..N...V..	.....V.
*08	.N.	VM.....	.....H	V..R..N...	.....	.....P..Q.	.....L.R	K..N.....	.....G.
*09	.F.	.....	.....L.V..	I..R.....	.....H.....	.....P..Q.	.....FV.L	KMSE.H.V..	.....G.
*10	.F.	.....	.....D..	.....Y...	.....	.....PS.Q.	.....FL.R	A..A.....	.....G.
*11	.F.	.....	.....D..	.....Y...	.....F.....	.....PS.Q.	.....FL.R	A..A.....	.....G.
*12	.Y.	.....	.....L..H	.....	.....F.....	.....	.....	.....	.....
*13	MEQ	LKG.....	.....L.V.F	I..R.....	.....LA..	.....PI.QN	.....V.R	T..E...V..	.....G.
*14	...	.....	.....	.....	.....	.....	.....	.....A.....	.....
*15	.F.	.....	.....D..	.....N...	.....	.....P..Q.	.....FV.L	KMSE.H.V..	.....GV
*16	MEQ	FKA.....	.....V.V	I..R..YA..	.....F.....	.....PA.Q.	L.....L.Q	M..N...V..	.....V.
				▲		*	*	*	*



