



# Evolution of MHC class I genes in Eurasian badgers, genus *Meles* (Carnivora, Mustelidae)

Shamshidin Abduriyim<sup>1</sup> · Yoshinori Nishita<sup>1,2</sup> · Pavel A. Kosintsev<sup>3</sup> · Evgeniy Raichev<sup>4</sup> · Risto Väinölä<sup>5</sup> · Alexey P. Kryukov<sup>6</sup> · Alexei V. Abramov<sup>7</sup> · Yayoi Kaneko<sup>8</sup> · Ryuichi Masuda<sup>1,2</sup>

Received: 6 April 2018 / Revised: 30 May 2018 / Accepted: 30 May 2018 / Published online: 29 June 2018  
© The Genetics Society 2018

## Abstract

Because of their role in immune defense against pathogens, major histocompatibility complex (MHC) genes are useful in evolutionary studies on how wild vertebrates adapt to their environments. We investigated the molecular evolution of MHC class I (MHCI) genes in four closely related species of Eurasian badgers, genus *Meles*. All four species of badgers showed similarly high variation in MHCI sequences compared to other Carnivora. We identified 7–21 putatively functional MHCI sequences in each of the badger species, and 2–7 sequences per individual, indicating the existence of 1–4 loci. MHCI exon 2 and 3 sequences encoding domains  $\alpha 1$  and  $\alpha 2$  exhibited different clade topologies in phylogenetic networks. Non-synonymous nucleotide substitutions at codons for antigen-binding sites exceeded synonymous substitutions for domain  $\alpha 1$  but not for domain  $\alpha 2$ , suggesting that the domains  $\alpha 1$  and  $\alpha 2$  likely had different evolutionary histories in these species. Positive selection and recombination seem to have shaped the variation in domain  $\alpha 2$ , whereas positive selection was dominant in shaping the variation in domain  $\alpha 1$ . In the separate phylogenetic analyses for exon 2, exon 3, and intron 2, each showed three clades of *Meles* alleles, with rampant trans-species polymorphism, indicative of the long-term maintenance of ancestral MHCI polymorphism by balancing selection.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41437-018-0100-3>) contains supplementary material, which is available to authorized users.

✉ Ryuichi Masuda  
masudary@sci.hokudai.ac.jp

- 1 Department of Natural History Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan
- 2 Department of Biological Sciences, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan
- 3 Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Ekaterinburg 620144, Russia
- 4 Agricultural Faculty, Trakia University, 6000 Stara Zagora, Bulgaria
- 5 Finnish Museum of Natural History, University of Helsinki, P. O. Box 17, FI-00014 Helsinki, Finland
- 6 Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok 690022, Russia
- 7 Zoological Institute, Russian Academy of Sciences, Saint Petersburg 199034, Russia
- 8 Carnivore Ecology and Conservation Research Group, Tokyo University of Agriculture and Technology, Tokyo 183-8509, Japan

## Introduction

The major histocompatibility complex class I (MHCI) proteins are expressed in all somatic cells (Bjorkman and Parham 1990), where they perform a variety of cellular tasks by presenting self-derived and foreign cytosolic peptides to receptors on CD8<sup>+</sup> cytotoxic T lymphocytes (Cresswell 2005). MHCI proteins thus play crucial roles in the vertebrate adaptive immune system by providing defense against various pathogens, and in autoimmunity. The MHCI protein complex comprises two non-covalently conjugated components: the MHCI heavy chain, which is encoded in the MHC region on chromosome 6 in humans, and on chromosome 12 or 18 in dogs (Xiao et al. 2016); a non-glycosylated light chain of  $\beta_2$ -microglobulin, which is a non-MHC protein ~100 amino acids long (see York and Rock 1996). The MHCI glycoprotein comprises a cytoplasmic tail, transmembrane domain, and extracellular  $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 3$  domains (Bjorkman and Parham 1990) encoded by exons 2, 3, and 4, respectively. The MHCI  $\alpha 1$  and  $\alpha 2$  domains are usually extensively polymorphic at both the nucleotide and amino-acid levels, especially for amino-acid residues that function as antigen-binding sites (ABS;

Hughes and Nei 1988), where MHCI proteins interact with intracellular peptides. Individuals and/or populations that are highly polymorphic for this region have a greater chance of responding to novel pathogens (Hughes and Nei 1992). The genetic diversity of MHC genes thus serves as an indicator of the immunological fitness of a population and its ability to respond to diseases (Sommer 2005). MHC also appears to be important in sexual selection (Sin et al. 2015; Gessner et al. 2017), kin recognition (Zelano and Edwards 2002), individual odors (Brown et al. 1989), and pregnancy outcomes (Kydd et al. 2016).

The diversity of MHC genes, particularly within exon 2 of MHC class II and exons 2 and 3 of MHCI, is presumably maintained by positive selection (see Akiyama et al. 2017; Saka et al. 2017), which fixes and accumulates beneficial point mutations through interactions between pathogens and their hosts. Its polymorphism is also retained by balancing selection (Nishita et al. 2015, 2017)—overdominance and frequency-dependent selection—which enhances the high allelic diversity within species and maintains ancestral allelic diversity over long periods. Some mammalian MHC allelic lineages have persisted for more than a million years, even across speciation events (Figueroa et al. 1988; Abduriyim et al. 2017). This trans-species polymorphism (TSP) can be detected in phylogenetic reconstructions (Nishita et al. 2015, 2017; Abduriyim et al. 2017; Akiyama et al. 2017; Maibach et al. 2017; Saka et al. 2017), where some alleles from one species are more closely related (sometimes even identical) to alleles from a different species than to alleles from the same species (Klein 1987; Klein et al. 2007). Phylogenetic analyses of MHCI genes can also show orthologous relationships, in which sequences are grouped by gene or locus, rather than by species (Gu and Nei 1999; Nei and Rooney 2005; Cao et al. 2015). The orthologous relationship is yet reflected in species closely related (Nei and Rooney 2005; Cao et al. 2015), because genes are diverged from their common ancestor by speciation.

Eurasian badgers, genus *Meles*, split into four distinct species—European badger, *M. meles*; Southwest Asian badger, *M. canescens*; Asian badger, *M. leucurus*; and Japanese badger, *M. anakuma* (Marmi et al. 2006; Sato 2016)—from a common ancestor more than 2 million years ago (Mya) (Marmi et al. 2006; Del Cerro et al. 2010); the last divergence of the Japanese badger from the continental Asian badger occurred between 0.21–1.09 Mya (Marmi et al. 2006). These species show differences in geographic variability, size, coat color and other morphological characteristics (Corbet 1978; Abramov 2003; Abramov and Puzachenko 2013; Sato 2016), and diet ecology (Kaneko et al. 2006; Li et al. 2013). They also differ in demographic histories, which might have affected the MHC diversity in the past: the Japanese badger revealed low genetic diversity

with possibly the most recent population growth (Kurose et al. 2001; Tashima et al. 2011), whereas European and Southwest Asian badgers underwent a historical sudden population expansion during the Middle Pleistocene (Marmi et al. 2006). These two badger species with a higher level of diversity have effective numbers of females six times greater than that of the Asian badger (Marmi et al. 2006). All four *Meles* species seem to have evolved under diverse pathogenic pressures, and some pathogens are species/region-specific in these species (Hancox 1980; Harasawa et al. 2014; Moreno et al. 2015; Hornok et al. 2017). A related finding is that MHCI genes in the European badger have been associated with the prevalence and/or intensity of certain pathogens (Sin et al. 2014). Moreover, these species possess several identical MHC class II *DRB* alleles, showing TSP at different taxonomic levels, and some alleles identified were novel to certain species and populations (Abduriyim et al. 2017). Comparative analyses revealed that the evolution of MHCI fundamentally differed from that of class II. Although MHC class II allelic lineages were shared, no such trans-species sharing of allelic lineages was seen at the MHCI loci in some primate species (Boyson et al. 1996), mostly due to the faster evolutionary rate of MHC class I than that of class II (Takahashi et al. 2000; Piontkivska and Nei 2003). Thus, all these facts present a compelling argument to analyze the variation and evolution of MHCI genes in the Eurasian badger. This would also contribute to insights into whole MHC diversity in these *Meles* species.

The MHC class II, but not class I, gene-assortative mating preference was evidenced in *M. meles* (Sin et al. 2015), although both MHC classes I and II genes have been associated with the prevalence and/or intensity of certain pathogens (Sin et al. 2014). Differences in the allelic diversity of MHC class II *DRB* genes have been detected among UK (Sin et al. 2012a) and continental populations of *M. meles*; 16 novel alleles were found from the continental population of *M. meles*, and the other 3 *Meles* species also showed higher genetic diversity (Abduriyim et al. 2017). Sin et al. (2012b) characterized MHCI genes in *M. meles* and found signs of separate evolutionary histories of MHCI protein domains: they studied insular population in UK and biased their phylogenetic analyses toward distantly related taxa. Thus, analyzing MHCI genes in closely related species is also needed to obtain further evidence for their findings.

In this study we characterized highly variable MHCI exons 2 and 3 and intervening intron 2 in four *Meles* species to compare MHCI gene content and diversity. Specifically, we addressed whether there are signatures of positive selection and recombination in *Meles* MHC alleles; how the different alleles are related among *Meles* species and to alleles in other carnivorans; and whether the *Meles* MHCI  $\alpha 1$  and  $\alpha 2$  domains evolved differently. Detailed



**Fig. 1** Map of Eurasia indicating locations where *Meles* samples analyzed in this study were obtained. Symbols: squares, *M. meles*; hexagons, *M. canescens*; circles, *M. leucurus*; stars, *M. anakuma*. Sample details are shown in Supplementary Table S1

comparative analyses of the MHC I gene diversity among the closely related *Meles* species can provide insights into understanding the evolutionary mechanisms driving variation in the immune system.

## Materials and methods

### Samples and DNA extraction

We used 25 samples from Eurasian badgers, of which 9 were from *M. meles*, 8 from *M. leucurus*, 6 from *M. anakuma*, and 2 from *M. canescens*. Sampling localities and profiles were as shown in Fig. 1 and Supplementary Table S1. Unfortunately, we could obtain only two individual samples from *M. canescens*, and the two samples may not fully represent the MHC gene characteristics in this species. Total DNA was extracted by using the DNeasy Blood and Tissue Kit (Qiagen), according to the manufacturer's protocol, and preserved in TE buffer at 4 °C until use.

### PCR amplification

Part of the MHC I gene spanning exons 2 and 3, and including the intervening intron 2, was amplified from the samples by PCR using the primers Meme-MHC Iex2F and Meme-MHC Iex3R (Sin et al. 2012a). PCR amplifications were conducted in 25  $\mu$ L volumes, each sample containing 7.5 pmol of phosphorylated primer, 125  $\mu$ M of dNTP, 1 $\times$  buffer ( $Mg^{2+}$ ), 0.62 U of PrimeSTAR GXL DNA polymerase (Takara, Japan), and 10–100 ng of genomic DNA. Cycling conditions in the Takara Dice Touch Thermal Cycler were 2 min at 94 °C; 27–30 cycles of 10 s at 98 °C, 15 s at 60 °C, and 30 s at 68 °C; and a final hold at 4 °C. The amplicons were subsequently separated by electrophoresis

on 2% agarose gels and visualized with ethidium-bromide fluorescence to confirm molecular size and quantity. Target amplicons were then purified using the QIAquick PCR Purification Kit (Qiagen).

### Obtaining and processing sequence data

PCR amplicons were cloned to obtain separate single sequences. Cloning and sequencing were done as described by Abduriyim et al. (2017). Following blue/white screening, 20–31 positive clones were sequenced for each individual using M13 forward and reverse primers. To determine bona fide MHC I sequences, we followed the criteria used for dog DLA (Kennedy et al. 2000). Identical sequences derived from at least two badger individuals or from two separate PCR reactions from the same individual were accepted as bona fide; unique single sequences were discarded as possible chimeras. After validation of the forward and reverse primer reads for each sequence, the vector and primer sequences were removed. The target sequences were aligned and the consensus sequences checked by using MEGA 6.0 (Tamura et al. 2013) and/or Seqman (Swindell and Plasterer 1997) in the DNASTAR Lasergene package. DnaSP v. 5 (Librado and Rozas 2009) was used to identify identical sequences among all sequences obtained from each badger species. To detect identical or homologous sequences in DNA databases, all unique MHC I sequences obtained from the four badger species were subjected to NCBI BLAST-nuc searches (Altschul et al. 1990). Verified MHC I sequences were named in compliance with the nomenclature conventions proposed by Klein et al. (1990). Verified sequences showing any signs of deletion, insertion, or premature stop codons within exons were regarded as pseudogene, and others as a presumably functional allele (PFA).

## Data analyses

Parameters related to variation were calculated for each species with MEGA 6.0 and DnaSP v.5. Non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitution rates were calculated with MEGA 6.0, using the method of Nei and Gojobori (1986) with the Jukes and Cantor (1969) correction. To assess historical selection pressure, the ratio  $d_N/d_S$  ( $\omega$ ) was calculated separately for ABS codons, determined according to Bjorkman et al. (1987), and non-ABS codons, separately for each species. Z-tests of historical positive selection were conducted in MEGA 6.0. Additionally, to search for evidence of positive selection site by site, a codon-based likelihood analysis was performed with CODEML implemented in PAML 4.4b (Yang 2007). To check for positively selected sites (PSS), which are indicated by  $\omega$  values significantly greater than 1, we used the following PAML models (Yang and Nielsen 2002; Yang et al. 2005): M1a (nearly neutral), M2a (positive selection), M7 (beta), and M8 (beta and  $\omega$ ). To determine whether the positive-selection models M2a and M8 provided a significantly better fit than the neutral-evolution models M1a and M7, we performed likelihood ratio tests (LRTs) to compare twice the difference in log likelihood using a  $\chi^2$  distribution. PSSs were identified by posterior probability values greater than 95% for models M2a and M8 using Bayes empirical Bayes inference, which is not affected by recombination events (Yang et al. 2005). Only sites where both the M2a and M8 models gave significant signatures were considered to be under positive selection.

Recombination analyses were implemented for the nucleotide alignments spanning exons 2 and 3, and for the intervening intron 2 in RDP4 (Martin et al. 2015). In the first run, we employed the RDP (Martin et al. 2010), GENECONV (Padidam et al. 1999), MaxChi (Smith 1992), and Bootscan (Martin et al. 2005) methods with default settings, using the Bonferroni correction for multiple comparisons to detect recombination events. Recombination events detected by at least three of these methods were then rechecked using all RDP methods available (Martin et al. 2010). All apparent recombination events detected in this manner were verified by MaxChi plots and matrices and neighbor-joining trees from inferred fragments, to confirm recombinant sequences and breakpoint placement. Only recombination events verified in this manner were regarded as having significant support. Finally, a genetic algorithm for recombination detection (GARD; Kosakovsky Pond et al. 2006), provided by the Datamonkey webserver (Delport et al. 2010), was also applied to inspect recombination signatures. While both RDP4 and GARD deduce the position of breakpoints, RDP4 also identifies highly likely recombinants and their parent sequences. Due to the similar effects of recombination and gene conversion on the

evolution of the <1000 bp sequence regions, they were not differentiated and are hence hereafter referred as recombination in the general sense (Richman et al. 2003).

We performed phylogenetic analyses separately for the nucleotide alignments of badger MHCI exon 2, intron 2, and exon 3 sequences, together with homologous sequences from some other carnivore species available from GenBank, including dog (*Canis lupus familiaris*) DLA, cat (*Felis catus*) FLA, ocelot (*Leopardus pardalis*), cheetah (*Acinonyx jubatus*), tiger (*Panthera tigris tigris*), and European badger (*Meles meles*). Sequences from *Lynx* species (Marmesat et al. 2017) and other carnivores were not included, due to shorter sequence lengths than those we obtained. All sequences were aligned by using MEGA 6.0 with the Gblocks option to yield the greatest number of informative nucleotide positions (Talavera and Castresana 2007). As a preferable alternative to phylogenetic trees in cases where gene duplication, recombination, and conversion likely occurred, neighbor-net networks for the MHCI sequences were reconstructed by using the uncorrected *P*-distance method implemented in Splitstree4 v4.14.5 (Huson and Bryant 2006). Nodal support was assessed by analysis of 1000 bootstrap pseudoreplicates, with only support values >70% being indicated in our trees.

## Results

### Diversity of MHCI sequences

A total of 588 clones were sequenced from 25 individuals representing four badger species, with a mean of 23.5 clones sequenced per individual. In BLAST-nuc searches, our sequences showed 95–100% similarity to sequences of *Meles* MHCI genes registered in GenBank. The final aligned MHCI data set comprised exon 2 (252 bp, 84 amino acids), exon 3 (255 bp, 85 amino acids), and intervening intron 2 (variable; 188–244 bp). We detected 64 distinct MCHI alleles among the four species, including 15 from *M. meles* (of which 12 were novel), 19 from *M. leucurus*, 7 from *M. canescens*, and 23 from *M. anakuma* (Supplementary Tables S2–S5). The deduced amino-acid sequences of PFAs all showed classical MHCI features. The highest number of PFAs detected in a single individual was six in *M. meles*, five in *M. leucurus*, seven in *M. canescens*, and six in *M. anakuma* (Supplementary Tables S2–S5), indicating that at least three loci exist in *M. leucurus*, *M. meles*, and *M. anakuma*, and four loci in *M. canescens*.

Based on inferred frame shifts or truncation in open reading frames, we identified presumed MHCI pseudogene sequences in *M. meles* and *M. anakuma*. *Meme-MHCI\*PS05* from *M. meles* showed a deletion of nucleotide positions 31–33, and *Meme-MHCI\*PS06* showed a two-



**Table 1** Sequence polymorphism in MHC1 exons 2 and 3, and intervening intron 2, for *M. anakuma*, *M. canescens*, *M. leucurus*, and *M. meles*

Species	<i>M. anakuma</i>			<i>M. leucurus</i>			<i>M. meles</i>			<i>M. canescens</i>		
	Exon 2	Exon 3	Intron 2	Exon 2	Exon 3	Intron 2	Exon 2	Exon 3	Intron 2	Exon 2	Exon 3	Intron 2
Variable sites	58	72	56	58	65	82	57	70	58	49	56	86
Parsim informative sites	49	59	49	52	60	76	44	42	28	23	33	55
Mutations	67	78	60	64	72	96	62	77	63	53	58	98
Synonymous	8	23	–	10	22	–	12	24	–	8	21	–
Non-synonymous	47	29	–	39	27	–	39	29	–	41	20	–
Number of amino acids	84	85	–	84	85	–	84	85	–	84	85	–
Polymorphic amino acid sites	34	32	–	30	30	–	32	30	–	29	25	–
Nucleotide diversity	0.079	0.086	0.098	0.083	0.102	0.199	0.078	0.094	0.113	0.083	0.098	0.223

– Indicates no values calculated because the intron is non-coding

nucleotide insertion between positions 113 and 114 in exon 2. *Mean-MHCI\*PS07* from *M. anakuma* showed a nucleotide insertion between positions 24 and 25 in exon 2. *Mean-MHCI\*PS01–05* from *M. anakuma* showed a premature stop codon at amino acid position 128 encoded from exon 3 (Supplementary Fig. S1). We included these presumed pseudogene sequences in phylogenetic analyses, but not in analyses of diversity, selection, and recombination. Among the 64 alleles from the four badger species there were 25 distinct intron-2 sequences (Supplementary Fig. S2), with length variants of 188 bp (number of sequences with intron-2 size of 188 bp among 64 alleles identified, 24), 189 bp (5), 199 bp (21), 244 bp (9), 215 bp (2), 194 bp (1), 195 bp (1), and 201 bp (1).

For all four *Meles* species, at the nucleotide level MHC1 exon 3 showed a higher number of variable sites than exon 2, but fewer non-synonymous substitutions; the number of polymorphic sites in the amino-acid sequences was similar between the two exons (Table 1). Intron 2 showed a higher number of variable sites than exons 2 and 3 in *M. leucurus* and *M. canescens*, but a similar number to exon 2 in *M. meles* and *M. anakuma* (Table 1). For all four species, the nucleotide diversity was lowest for exon 2 (0.079–0.083), intermediate for exon 3 (0.086–0.102), and highest for intron 2 (0.098–0.223; Table 1).

### Selection and recombination analyses

We analyzed selection pressures separately for domains  $\alpha 1$  (encoded by exon 2) and  $\alpha 2$  (exon 3), as their evolutionary histories might have been different. The non-synonymous substitution rate ( $d_N$ ) was higher than the synonymous rate ( $d_S$ ) for ABSs in both domains in all four species. However,  $\omega$  ( $d_N/d_S$ ) values were  $>1$  for ABSs in the  $\alpha 1$  domain, but not in the  $\alpha 2$  domain in all species, indicating that, historically, positive selection to maintain allelic diversity was stronger in the  $\alpha 1$  domain than in the  $\alpha 2$  domain (Table 2). PAML positive selection models M2a and M8 showed a

better fit than did neutral evolution models M1a and M7 in all four species, except for the  $\alpha 1$  domain in *M. canescens* MHC1 (Table 3). Features of positive selection were found in both domains  $\alpha 1$  and  $\alpha 2$ . Parameter estimates under the M2a and M8 models indicated that more sites (35–37%) were under positive selection in the  $\alpha 1$  domain than in the  $\alpha 2$  domain (5–6%) in *M. leucurus*, *M. canescens*, and *M. anakuma*. A few sites (about 4%) were under positive selection in both domains in *M. meles* (Table 3). Two to five sites were identified as under positive selection, and PSSs invariably fell within ABSs in all four badger species (Supplementary Fig. S1).

RDP4 detected one and three recombination events in *M. anakuma* and *M. meles* MHC1 sequences, respectively (Supplementary Table S6). Two recombination breakpoints (RBPs) were detected for *M. anakuma* (2 RBPs), *M. meles* (6 RBPs) in RDP4. GARD, on the other hand, detected recombination signals in *M. leucurus* as well as in *M. meles* and *M. anakuma*, but not in *M. canescens*. The RBPs detected by GARD did not exactly correspond to those detected by RDP4 (Supplementary Table S6), although the positioning of RBPs in the exons and introns was similar.

### Phylogenetic analyses

In the phylogenetic reconstruction for MHC1 exon 2 (Fig. 2 and Supplementary Fig. S3a), *Meles* sequences did not group monophyletically by species, but formed three clades (A, B, and C), with each clade containing sequences from all four species. TSP was clearly evident, with some exon 2 sequences from particular *Meles* species being more closely related to sequences from other *Meles* species than to those from the same species. Clade B formed the sister group to a clade containing sequences from the giant panda, and in clade A seven exon 2 sequences from the giant panda formed a basal clade.

Three clades containing *Meles* sequences were also evident in the phylogenetic reconstruction for exon 3

**Table 2** Rates of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions and the ratio  $\omega$  ( $d_N/d_S$ ) for antigen-binding sites (ABS) codons, non-ABS codons, and all codons (combined) in MHCI exons 2 and 3 for *M. anakuma*, *M. canescens*, *M. leucurus*, and *M. meles*

	<i>M. anakuma</i>			<i>M. leucurus</i>			<i>M. meles</i>			<i>M. canescens</i>		
	ABS	Non-ABS	Combined	ABS	Non-ABS	Combined	ABS	Non-ABS	Combined	ABS	Non-ABS	Combined
Exon 2												
$d_S$	0.11 ± 0.08	0.29 ± 0.19	0.23 ± 0.15	0.03 ± 0.02	0.29 ± 0.10	0.24 ± 0.09	0.43 ± 0.11	0.42 ± 0.15	0.42 ± 0.14	0.17 ± 0.24	0.21 ± 0.53	0.24 ± 0.19
$d_N$	1.13 ± 0.14	0.28 ± 0.08	0.43 ± 0.10	0.91 ± 0.09	0.27 ± 0.06	0.40 ± 0.08	1.24 ± 0.09	0.31 ± 0.06	0.48 ± 0.09	0.54 ± 0.26	0.17 ± 0.16	0.05 ± 0.47
$\omega$	9.87**	0.96	1.90**	26.38**	0.91	1.67	2.90**	0.74	1.14*	3.24**	0.81	0.22
Exon 3												
$d_S$	0.85 ± 0.27	0.46 ± 0.14	0.57 ± 0.16	1.66 ± 0.39	0.42 ± 0.10	0.65 ± 0.18	1.21 ± 0.30	0.58 ± 0.12	0.65 ± 0.14	0.79 ± 0.72	0.32 ± 0.31	0.39 ± 0.37
$d_N$	1.05 ± 0.19	0.21 ± 0.06	0.38 ± 0.11	1.30 ± 0.18	0.18 ± 0.04	0.38 ± 0.10	0.94 ± 0.12	0.23 ± 0.05	0.38 ± 0.08	0.62 ± 0.38	0.12 ± 0.12	0.22 ± 0.21
$\omega$	1.22	0.46	0.66	0.78	0.42	0.59	0.78	0.39	0.59	0.78	0.39	0.56

\* Statistically significant ( $P < 0.05$ ), \*\* statistically significant ( $P < 0.01$ )

(Supplementary Fig. S3b). Although feline, canine, ursine, and *Meles* sequences tended to form separate clades, a few sequences (such as Acju-AJUI3 and L.pardalis\_LPAI69K) nonetheless showed TSP within the feline group, and two canid sequences fell outside the canine group, positioned basal to *Meles* cluster C.

The phylogenetic reconstruction for intron 2 sequences likewise showed distinct *Meles* clades A, B, and C (Supplementary Fig. S3c). These clades included exactly the same allele sequences for intron 2 as for exon 3, except that *Meme-MHCI\*1PS06* and *Mele-MHCI\*02* were in clade A for intron 2 but in clade B for exon 3.

## Discussion

### Diversity of *Meles* MHCI genes

Kaufman et al. (1994) delineated well-conserved amino-acid residues related to structural features characteristic of classical MHCI proteins. For instance, a cysteine (C) residue forms an intradomain disulfide bridge in the  $\alpha 2$  domain; other characteristic residues include an N-linked glycosylation site in  $\alpha 1$ , a threonine (T) to interact with the TAP (transporter associated with antigen processing) complex, residues in  $\alpha 3$  that bind to CD8 glycoprotein on the surface of T cells, and eight highly conserved residues in  $\alpha 1$  and  $\alpha 2$  in mammals that bind to the N-termini and C-termini of peptides. Sin et al. (2012b) found all these features in a transcriptional analysis of the UK *M. meles* population. As expected, within the deduced amino-acid sequences from PFAs we observed the same classical features of MHCI proteins in  $\alpha 1$  and  $\alpha 2$  (Supplementary Fig. S1). In addition, we detected extensive polymorphism for both  $\alpha 1$  and  $\alpha 2$  (Table 1), in contrast to monomorphic or oligomorphic non-classical MHCI molecules (Shawar et al. 1994). Taken together, these features demonstrate that most of the PFA sequences we obtained from all four *Meles* species are classical MHC class Ia genes.

The domestic dog has one classical MHCI gene with 73 alleles (Wagner et al. 2002; Ross et al. 2012; Venkataraman et al. 2017); domestic cat has three classical MHCI genes (Yuhki et al. 2008) with a number of alleles; and giant panda has four classical MHCI genes, of which Aime-I has 16 alleles (Pan et al. 2008; Zhu et al. 2013). Liu et al. (2017) reported 17 functional alleles from four MHCI genes in four wolf (*Canis lupus*) individuals. Therefore, it is not surprising that we detected 7–19 PFAs in the four *Meles* species (Supplementary Tables S2–S5). Our

**Table 3** Positively selected sites (PSSs), log-likelihood (lnL) values, and parameter estimates under four PAML models of codon evolution, and values of the test statistic (TS) and probability (*P*) for the likelihood ratio test (LRT) for MHC1 domains  $\alpha 1$  and  $\alpha 2$  in four Eurasian badger species

Species	Domain	Model	lnL value	Parameter estimates	PSSs	LRT	TS value	<i>P</i> -value
<i>M. anakama</i>	$\alpha 1$	M1a	-783.19	$P_0 = 0.52, P_1 = 0.48, \omega_0 = 0.00, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	8.18	<0.05
		M2a	-779.11	$P_0 = 0.56, P_1 = 0.09, P_2 = 0.35, \omega_0 = 0.03, \omega_1 = 1.00, \omega_2 = 2.78$	24S, 66I, 67C, 80T, 81L			
		M7	-783.26	$P = 0.01, q = 0.01$	Not allowed	M7 vs. M8	8.35	<0.05
		M8	-779.09	$P_0 = 0.63, P_1 = 0.11, q = 0.80, P_1 = 0.37, \omega = 2.72$	23I, 24S, 63E, 66I, 67C, 69E, 70T, 71T, 73T, 75R, 76E, 79N, 80I, 81L, 83D			
	$\alpha 2$	M1a	-802.91	$P_0 = 0.84, P_1 = 0.17, \omega_0 = 0.09, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	27.82	<0.001
		M2a	-789.00	$P_0 = 0.75, P_1 = 0.21, P_2 = 0.05, \omega_0 = 0.11, \omega_1 = 1.00, \omega_2 = 12.41$	97T, 114S, 116V, 156Y			
		M7	-803.75	$P = 0.19, q = 0.56$	Not allowed	M7 vs. M8	33.90	<0.001
		M8	-786.80	$P_0 = 0.95, P_1 = 0.51, q = 1.88, P_1 = 0.05, \omega = 10.04$	97T, 114S, 116V, 156Y			
<i>M. leucurus</i>	$\alpha 1$	M1a	-773.12	$P_0 = 0.59, P_1 = 0.41, \omega_0 = 0.00, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	6.83	<0.05
		M2a	-769.70	$P_0 = 0.63, P_1 = 0.00, P_2 = 0.37, \omega_0 = 0.02, \omega_1 = 1.00, \omega_2 = 2.16$	63Q, 67C			
		M7	-773.12	$P = 0.01, q = 0.01$	Not allowed	M7 vs. M8	6.84	<0.05
		M8	-769.70	$P_0 = 0.63, P_1 = 0.02, q = 0.47, P_1 = 0.37, \omega = 2.18$	24A, 43R, 45M, 63Q, 66I, 67C, 70A, 71A, 76E, 83G			
	$\alpha 2$	M1a	-843.05	$P_0 = 0.85, P_1 = 0.15, \omega_0 = 0.06, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	31.48	<0.001
		M2a	-827.31	$P_0 = 0.84, P_1 = 0.10, P_2 = 0.06, \omega_0 = 0.10, \omega_1 = 1.00, \omega_2 = 7.71$	97N, 114R, 116V, 156W			
		M7	-847.19	$P = 0.16, q = 0.52$	Not allowed	M7 vs. M8	41.16	<0.001
		M8	-826.61	$P_0 = 0.94, P_1 = 0.48, q = 2.25, P_1 = 0.06, \omega = 7.19$	97N, 114R, 116V, 152V, 156W			
<i>M. meles</i>	$\alpha 1$	M1a	-824.07	$P_0 = 0.57, P_1 = 0.43, \omega_0 = 0.00, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	20.88	<0.001
		M2a	-813.63	$P_0 = 0.55, P_1 = 0.41, P_2 = 0.03, \omega_0 = 0.00, \omega_1 = 1.00, \omega_2 = 8.75$	66N, 67L			
		M7	-824.18	$P = 0.01, q = 0.01$	Not allowed	M7 vs. M8	20.94	<0.001
		M8	-813.71	$P_0 = 0.96, P_1 = 0.01, q = 0.01, P_1 = 0.04, \omega = 8.60$	52M, 66N, 67L, 73A, 79N, 81L			
	$\alpha 2$	M1a	-809.55	$P_0 = 0.83, P_1 = 0.17, \omega_0 = 0.07, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	13.16	<0.01
		M2a	-802.97	$P_0 = 0.81, P_1 = 0.15, P_2 = 0.04, \omega_0 = 0.08, \omega_1 = 1.00, \omega_2 = 5.39$	114S, 116V			
		M7	-810.35	$P = 0.16, q = 0.55$	Not allowed	M7 vs. M8	18.31	<0.001
		M8	-801.20	$P_0 = 0.94, P_1 = 0.3, q = 1.82, P_1 = 0.06, \omega = 3.98$	114S, 116V			
<i>M. canescens</i>	$\alpha 1$	M1a	-581.90	$P_0 = 0.49, P_1 = 0.51, \omega_0 = 0.00, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	2.63	>0.05
		M2a	-580.59	$P_0 = 0.63, P_1 = 0.00, P_2 = 0.37, \omega_0 = 0.10, \omega_1 = 1.00, \omega_2 = 2.22$	NS			
		M7	-581.90	$P = 0.01, q = 0.01$	Not allowed	M7 vs. M8	2.63	>0.05
		M8	-580.59	$P_0 = 0.63, P_1 = 11.25, q = 99.00, P_1 = 0.37, \omega = 2.22$	23I, 24A, 72P, 66I, 67C, 70A, 81A			
	$\alpha 2$	M1a	-600.27	$P_0 = 0.88, P_1 = 0.12, \omega_0 = 0.06, \omega_1 = 1.00$	Not allowed	M1a vs. M2a	10.95	<0.01
		M2a	-594.80	$P_0 = 0.94, P_1 = 0.00, P_2 = 0.06, \omega_0 = 0.11, \omega_1 = 1.00, \omega_2 = 5.69$	116D, 156W			
		M7	-602.64	$P = 0.15, q = 0.58$	Not allowed	M7 vs. M8	15.71	<0.001
		M8	-594.79	$P_0 = 0.93, P_1 = 5.07, q = 41.30, P_1 = 0.06, \omega = 5.78$	114S, 116D, 156W			

$\omega$  ratio of non-synonymous to synonymous substitution rates ( $d_N/d_S$ ),  $P_n$  proportion of amino acids in the  $\omega_n$  site class,  $P$  and  $q$  parameters of the beta distribution. The percentages of PSSs in the domains are underlined

Values for lnL,  $\omega$ ,  $P_n$ ,  $\omega_n$ ,  $P$ , and  $q$  were calculated with CODEML in PAML 4

TS is twice the difference in lnL values from alternative models vs. their null models; d.f. = 2 for all LRTs; *P*-values were obtained by comparison of TS to a  $\chi^2$  distribution using CHI2 in the PAML 4 package

PSSs were identified in models M2a and M8 by Bayes Empirical Bayes (BEB), with posterior probabilities (PP) > 95%; codons with PP > 99% are in bold font; NS not significant

findings of allele numbers would indicate the possibility of detecting more novel alleles in other populations of *Meles* species.

PFAs detected per individual (Supplementary Tables S2–S5) indicated that *M. canescens* could have more loci than the other three badger species, but this needs corroboration, as we studied limited number of samples from *M. canescens*. Among a thousand individuals in the UK population of *M. meles*, Sin et al. (2012b, 2014, 2015) detected seven PFAs, indicating the existence of at least two loci. We identified 13 PFAs (Supplementary Table S2), which might be from 1–3 loci, from only nine individuals of continental *M. meles*. These results parallel those for MHC class II *DRB* genes in UK and continental *M. meles* (Sin et al. 2012a; Abduriyim et al. 2017), suggesting that the UK population shows lower MHC allelic diversity than the continental populations. The UK population of *M. meles* reported to have more alleles of *MHCI* gene than those of MHC *DRB* (Sin et al. 2012a, b, 2014, 2015); the reverse was observed in its counterparts in the continent, and in the *M. canescens*, *M. leucurus*, and *M. anakuma* species (Abduriyim et al. 2017; Supplementary Tables S2–S5), indicating that they might be subject to a different level and/or type of selection pressure from pathogens or to some effect of genetic drift in a small insular population. Intriguingly, we detected allele *Meme-MHCI\*05* in five of nine continental *M. meles* individuals, and the allele *Mele-MHCI\*15*, which has identical nucleotide sequences to *Meme-MHCI\*05* in exons 2 and 3 in three of eight *M. leucurus* individuals. In addition, exons 2 and 3 in three alleles of *M. meles*, *Meme-MHCI\*05*, *\*12*, and *\*14*, encode an identical amino-acid sequence; therefore, seven of the nine *M. meles* individuals have identical partial amino-acid sequences of *MHCI* protein, and identical sequences were also found in individuals of all *M. leucurus* examined (encoded by *Mele-MHCI\*04*, *\*06*, *\*07*, *\*10*, and *\*12*), in all *M. canescens* (*Meca-MHCI\*02*, *\*03*, and *\*06*), and in two of six *M. anakuma* (*Man-MHCI\*09* and *\*12*) (Table 4). *Meme-MHCI\*05* was as well identified in thousands of individuals studied in the UK population (Sin et al. 2014); thus, it appears that the amino-acid sequence deduced from *Meme-MHC\*05* is the dominant sequence and has an important function in *Meles* species.

There are two types of *MHCI* pseudogenes: complete genes whose expression is interrupted by frame shifts, premature stop codons, or other defects; and gene fragments lacking one or more exons characteristic of expressed genes (Hughes 1995). Both pseudogenes and novel functional genes are thought to have originated from classical *MHCI* loci by gene duplication (Hughes 1995; Beck et al. 1999). We identified several presumed pseudogene sequences with mutations leading to frame shifts or premature stop codons in *M. meles* and *M. anakuma*, but not in *M. leucurus* or *M.*

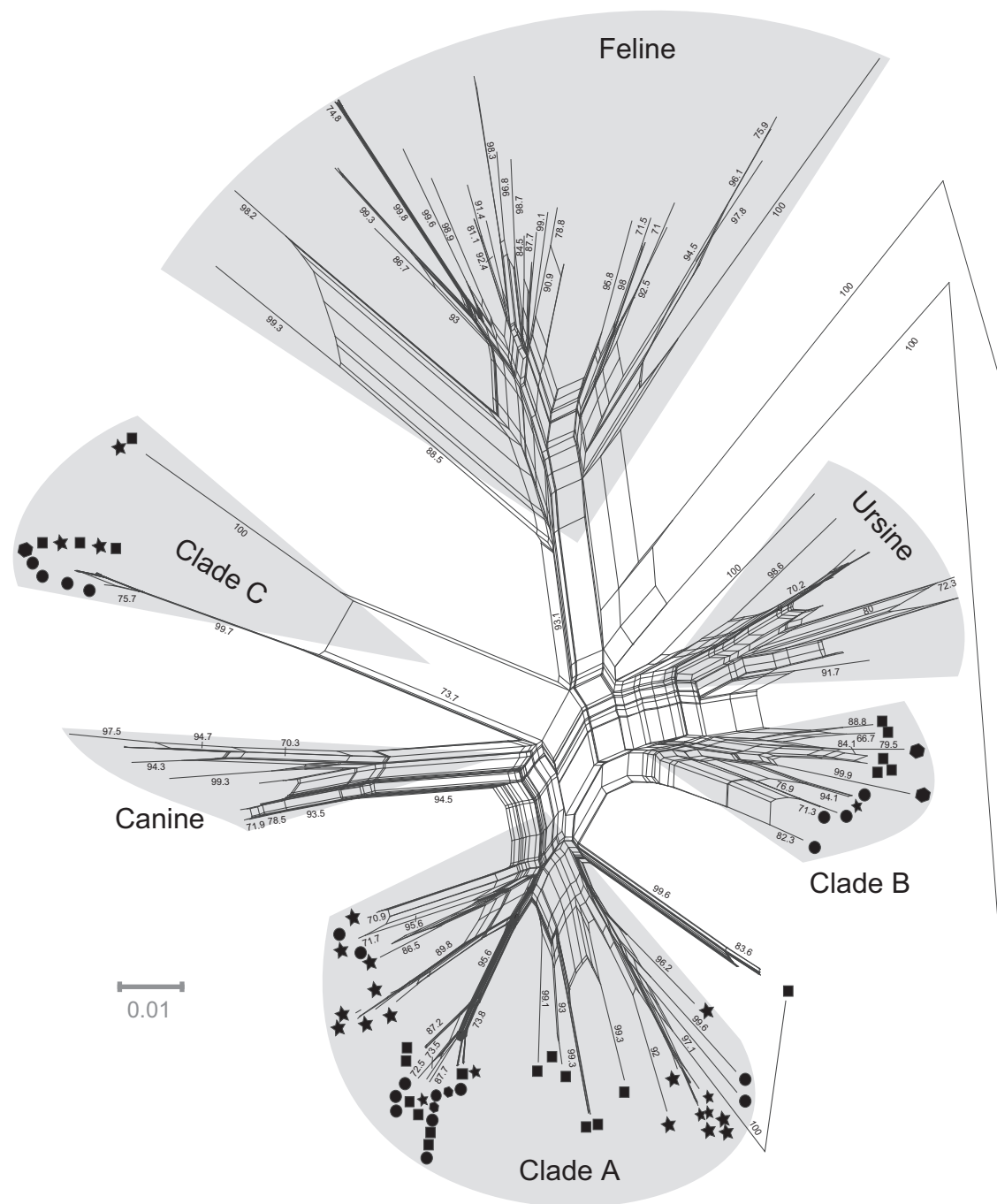
*canescens*. The pseudogenes varied in frequency; for example, pseudogene *Meme-MHCI\*PS05* with a deletion was detected in four of nine *M. meles* individuals, while *Meme-MHCI\*PS06* with an insertion was found in only one individual. Each five of six *M. anakuma* individuals contained at least two pseudogenes.

The locus-specific intron sequences flanking the MHC polymorphic exons may prove useful for PCR-based MHC typing strategies (Cereb et al. 1995). The intron 2 of *Meles* sequences (Supplementary Fig. S2) revealed that most of the mutations are indels (insertions and/or deletions), leading to size variations of 188–244 bp. Shorter intron 2 sequences were documented for *M. meles* (Sin et al. 2012b), while longer intron 2 sequences were found in pseudogenes in canids and *M. meles* (Wagner 2003; Sin et al. 2012b). However, longer intron 2 of functional genes was recorded in many other mammal species (Cereb et al. 1996). In addition, supporting our results, the feline intron 2 sequences likewise varied between 182 and 202 bp (Pokorny et al. 2010; Castro-Prieto et al. 2011).

## Evolution of *Meles* *MHCI* genes

We found no evidence of historical positive selection based on accumulated non-synonymous substitutions for ABS or non-ABS codons in  $\alpha 2$  domain in any of the target badger species (Table 2), which parallels the results for the UK population (Sin et al. 2012b). The PAML analysis, however, identified particular codons under positive selection in both domains  $\alpha 1$  and  $\alpha 2$  (Table 3), where all PSSs fell within ABSs (Supplementary Fig. S1). It is conceivable to detect positive selection in both these domains, as domains  $\alpha 1$  and  $\alpha 2$  are both responsible for peptide binding (Kaufman et al. 1994). Our results of positive selection analysis using PAML are inconsistent with the study by Sin et al. (2012b), which detected positive selection only in domain  $\alpha 1$  but not in domain  $\alpha 2$  for the UK population of *M. meles*. In contrast, our study indicated that both domains are under positive selection pressure, but the intensity of positive selection was likely quite different for domains  $\alpha 1$  and  $\alpha 2$  across all four *Meles* species, except for those from UK and continental populations of *M. meles*. In turn, positive selection shaping and maintaining polymorphisms in ABS can lead to a wide range of antigen peptides in any given species or population, and may increase the chance of survival against pathogenic and parasitic infections (Hughes and Nei 1992; Hughes and Yeager 1998). Therefore, considering species/region-specific as well as common pathogens (Hancox 1980; Harasawa et al. 2014; Moreno et al. 2015; Hornok et al. 2017), we hypothesize that, in the continental population, the  $\alpha 2$  domain plays a crucial role in recognizing pathogens that are unfamiliar or novel to the UK population in *M. meles*, and that not only  $\alpha 1$  but also  $\alpha 2$





**Fig. 2** Phylogenetic neighbor-net for MHC I exon 2 (simplified) sequences from four *Meles* species from this study and including canine, feline, and ursine sequences obtained from GenBank.

Symbols: squares, *M. meles*; hexagons, *M. canescens*; circles, *M. leucurus*; stars, *M. anakuma*. Bootstrap support values of 70% or greater are shown. See detailed tree(s) in Supplementary Fig. S3

domain is important for pathogen recognition in the other three *Meles* species.

In the phylogenetic neighbor net, exon 2 and 3, and intron 2, sequences did not form species-specific clades among *Meles* species, but for each gene region distantly related clades A, B, and C contained sequences from more than one species, a common characteristic of MHC genes. TSP (Klein et al. 1998) is indicative of balancing selection,

resulting in ancestral allelic lineages being maintained for long periods, even after speciation events and subsequent diversification (Bernatchez and Landry 2003). The pattern of sequences clustering across species is ultimately due to orthology, which results in sequences clustering by locus (Gu and Nei 1999; Cao et al. 2015) rather than by species. MHC I intron 2 sequences tend to form locus-specific clusters in closely related species (Groot et al. 2002), which

**Table 4** Identical sequences of MHC I alleles between two or more of the *Meles* species included in the present study

Gene regions	<i>M. meles</i>			<i>M. anakuma</i>			<i>M. canescens</i>			<i>M. leucurus</i>		
	Ex 2	In 2	Ex 3	Ex 2	In 2	Ex 3	Ex 2	In 2	Ex 3	Ex 2	In 2	Ex 3
Allele	<i>Meme*14</i>						<i>Meca*02</i>					
Nucleic acid	■	■	■				■	■	■			
Amino acid	◎		◎				◎		◎			
Allele	<i>Meme*05</i>						<i>Mele*15</i>					
Nucleic acid	■	■	■							■	■	■
Amino acid	◎		◎							◎		◎
Allele	<i>Mean*09</i>						<i>Meca*03</i>					
Nucleic acid				■	■	■	■	■	■			
Amino acid				◎		◎	◎		◎			
Allele	<i>Mean*12</i>						<i>Meca*06</i>					
Nucleic acid				■	■	■	■	■	■			
Amino acid				◎		◎	◎		◎			
Allele	<i>Mean*13</i>						<i>Mele*14</i>					
Nucleic acid				■	■	■				■		■
Amino acid				○		○				○		○
Allele	<i>Meme*20</i>						<i>Meca*07</i>					
Nucleic acid	■	■	■				■	■	■			
Amino acid	○		○				○		○			○
Allele	<i>Meme*21</i>						<i>Mean*16</i>					
Nucleic acid	■		■	■		■						
Amino acid	○		○	○		○						
Allele	<i>Meme*12</i>						<i>Mele*04,06,07,10,12</i>					
Nucleic acid												
Amino acid	○		○							○		○

Ex and In signify exon and intron, respectively

A square (■) indicates identical nucleotide sequences among the aligned alleles

A circle (○) indicates identical amino-acid sequences among the aligned alleles

A double circle (◎) indicates identical amino-acid sequences among the whole alleles in this table

may be reflected in the high similarity in the composition of cluster C between intron 2 (Supplementary Fig. S3c) and exons 2 and 3 (Fig. 2, Supplementary Fig. S3a, b). In some cases, badger individuals had more than two alleles within a cluster in the phylogenetic tree, which indicates that orthologous loci underwent independent duplications in some species. Miska et al. (2002) argued that, with large-enough divergence times between species, orthologous relationships between loci could be lost. The divergence time between *Meles* species is more than 2 million years (Marmi et al. 2006; Del Cerro et al. 2010). Thus, this amount of evolutionary time is more likely insufficient to have eliminated orthologous relationships between loci in *Meles* species.

Recombination between loci can contribute to high allelic diversity in MHC genes (McAdam et al. 1994; Schaschl et al. 2005), and exon exchange between loci via recombination has been documented for mammalian MHC

(Hughes and Nei 1989; Gu and Nei 1999). In our study, both the RDP4 and GARD analyses detected signatures of recombination events in *M. anakuma* and *M. meles*, but only GARD detected recombination in *M. leucurus* (Supplementary Table S6). RBPs inferred by RDP4 did not exactly correspond to those inferred by GARD. However, these two methods often give inconsistent RBPs (Zhao et al. 2013; Huang et al. 2016), probably because they use different computational methods (Kosakovskiy et al. 2006; Martin et al. 2010). Detection of recombination is a complicated task, and hence improved methods are needed for the purpose (Bay and Bielawski 2011).

In our analyses, RDP4 identified RBPs in intron 2 and exon 3 in both *M. anakuma* and *M. meles*, while GARD identified RBPs in intron 2 and exon 3 in these two species and *M. leucurus*. Although these RBPs appear to be random, suggesting they might be PCR artifacts, PCR-mediated recombinant sequences will likely be present in

only a few clones and are unlikely to be sequenced twice (Huang et al. 2016), whereas the recombinant sequences we detected were each found in 3–19 clones. Partly corroborating this, we detected natural recombination events: the degenerate 13-bp sequence motif CCNCCNTNNCCNC, which is associated with human recombination hotspots (Myers et al. 2008), occurred in nearly half of the *Meles* intron 2 sequences, where most of the RBPs were located. No recombination hotspots are known for PCR-mediated recombinants (Cronn et al. 2002; Lahr and Katz 2009; Huang et al. 2016), indicating that the recombination events we detected did not occur in vitro, but in vivo. Sin et al. (2012b) likewise reported in vivo recombination events in the UK population of *M. meles*.

Incongruity in within-clade topology and positive selection intensity between exons 2 and 3 indicates that domains  $\alpha 1$  and  $\alpha 2$  had different evolutionary histories. This is to some degree in agreement with the findings of Sin et al. (2012b). The birth-and-death evolution model assumes that novel genes arise by gene duplication, and that some of the duplicated genes, which may represent ancestral polymorphism, are retained for long periods, while others are deleted or become non-functional (Nei and Rooney 2005). Because our study and Sin et al. (2012b) detected a number of pseudogenes derived from *Meles* MHCI PFA, and our phylogenetic analysis also indicated independent duplication of orthologous loci, it would be expected under the birth-and-death model of evolution. More recombinants with RBPs in intron 2 and exon 3 were found in *M. meles* than in any of the other species, corroborating the suggestion by Sin et al. (2012b) that the  $\alpha 2$  domain underwent a concerted evolution, in which RBPs in intron 2 and exon 3 separated domain  $\alpha 1$  from domain  $\alpha 2$  and/or  $\alpha 3$ , with the  $\alpha 2/\alpha 3$  domain sequences becoming homogenized. Although, Gu and Nei (1999) and Nei and Rooney (2005) argued that recombination is relatively rare and tends not to be selectively advantageous, and that variation in MHC genes is generated primarily by mutation and selection, we cannot rule out the proposal (McAdam et al. 1994; Schaschl et al. 2005) that mutant alleles were generated by recombination or gene conversion, as many recent studies, in addition to ours (Nishita et al. 2015, 2017; Abduriyim et al. 2017) and Sin et al. (2012b), have collectively reported recombination events in MHC.

We detected many more PSSs at ABSs in the  $\alpha 1$  domain than in the  $\alpha 2$  domain in *M. anakuma*, *M. leucurus*, and *M. canescens*, but equivalently low numbers of PSSs in domains  $\alpha 1$  and  $\alpha 2$  in *M. meles* (Table 3). The RDP4 and GARD analyses both detected recombination events in *M. anakuma* and *M. meles*, whereas no recombination signatures were evident in *M. canescens*, and only GRDG showed RBPs in *M. leucurus* (Supplementary Table S6). Most RBPs were in intron 2, with a few in exon 3.

Intriguingly, all four *Meles* species showed an equally high level of polymorphism for both domains  $\alpha 1$  and  $\alpha 2$  (Table 1). Taken together, these findings indicate that positive selection dominated in shaping the high polymorphism in  $\alpha 1$  domain, whereas both recombination and positive selection contributed to the high polymorphism in  $\alpha 2$  domain. Meanwhile, the possibilities obtained from the positive selection, diversity, and recombination analysis together suggest that different mechanisms shaped MHCI gene polymorphisms in these closely related species. This could be explained by the different biological traits, contemporary and historical demography, and/or the varying spectra of pathogens to which these species exposed. To comprehensively understand the complexity of MHC evolution in these *Meles* species, further pooled analysis on the genetic diversity of MHC and additional studies on demographic, biological and ecological traits, and pathogens is needed.

## Conclusion

Our findings emphasize the significance of exploring each component of a gene separately (Sin et al. 2012b) to gain comprehensive knowledge of MHCI evolution. Both MHC class I and II (Abduriyim et al. 2017) genes showed extensive TSP, which is indirect evidence of balancing selection. However, the separate phylogenetic analyses exhibited incongruity between domains  $\alpha 1$  and  $\alpha 2$  in MHCI proteins, implying that they have different evolutionary histories. As Sin et al. (2012b) proposed, the more rational interpretation could be a concerted evolution of the  $\alpha 2$  domain, given that RBPs in intron 2 probably lead to homogenizing exon 3 ( $\alpha 2$  domain), while exon 2 ( $\alpha 1$  domain) could be subject to balancing selection. A smaller number of shared alleles and higher number of pseudogenes were observed in MHCI compared to class II (Abduriyim et al. 2017). This is due to the faster turnover rate of typical MHCI loci relative to that of MHC class II. Owing to various degrees of divergence, MHCI classical genes generate a number of non-classical genes and pseudogenes (Piontkivska and Nei 2003).

The differences in selective pressures and recombination events, but equally high sequence variations in the four *Meles* species, indicate different mechanisms for shaping MHCI gene polymorphisms. The varying spectrum of pathogens, particularly species/region-specific pathogens, provides a more plausible explanation to drive and retain variations. Taking MHCI-associated autoimmunity and recombination events in these species into consideration, it can be concluded that the ethological, ecological, or biological traits and demographic histories, which are scarcely known, may to some degree play a role in MHC evolution.

## Data archiving

Sequence data have been submitted to GenBank: accession numbers LC350023–LC350083.

**Acknowledgements** We thank Dr. M Baba (Kitakyushu Museum of Natural History and Human History), T Tsujimoto (Morioka Zoo), S Dakemoto, and S Yachimori (Shikoku Institute of Natural History), and the Museum at the Institute of Plant and Animal Ecology (Ural Branch of the Russian Academy of Sciences) for providing samples and support in sampling, and Dr. M Dick for his helpful comments and for editing our English. We are especially grateful to three anonymous reviewers for their valuable comments and suggestions on our manuscript. This study was funded in part by a Joint Research Project Grant from the Japan Society of the Promotion of Science (JSPS) and the Russian Foundation for Basic Research (RFBR, grant 16-54-50004), and a grant from the Joint Research Program of the Japan Arctic Research Network Center.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abduriyim S, Nishita Y, Kosintsev PA, Raichev E, Väinölä R, Kryukov AP et al. (2017) Diversity and evolution of MHC class II DRB gene in the Eurasian badger genus *Meles* (Mammalia: Mustelidae). *Biol J Linn Soc* 122:258–273
- Abramov AV (2003) The head colour pattern of the Eurasian badgers (Mustelidae, *Meles*). *Small Carniv Conserv* 29:5–7
- Abramov AV, Puzachenko AY (2013) The taxonomic status of badgers (Mammalia, Mustelidae) from Southwest Asia based on cranial morphometrics, with the redescription of *Meles canescens*. *Zootaxa* 3681:44–58
- Akiyama T, Kohyama TI, Nishida C, Onuma M, Momose K, Masuda R (2017) Genetic variation of major histocompatibility complex genes in the endangered red-crowned crane. *Immunogenetics* 69:451–462
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Bay RA, Bielawski JP (2011) Recombination detection under evolutionary scenarios relevant to functional divergence. *J Mol Evol* 73:273–286
- Beck S, Geraghty D, Inoko H, Rowen L, Aguado B, Bahram S et al. (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401:921–923
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* 16:363–377
- Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem* 59:253–288
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506–512
- Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI (1996) The MHC class I genes of the rhesus monkey: different evolutionary histories of MHC class I and II genes in primates. *J Immunol* 156:4656–4665
- Brown RE, Roser B, Singh PB (1989) Class I and class II regions of the major histocompatibility complex both contribute to individual odors in congenic inbred strains of rats. *Behav Genet* 19:659–674
- Cao YH, Fan JW, Li AX, Liu HF, Li LR, Zhang CL et al. (2015) Identification of MHC I class genes in two Platyrrhini species. *Am J Primatol* 77:527–534
- Castro-Prieto A, Wachter B, Sommer S (2011) Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Mol Biol Evol* 28:1455–1468
- Cereb N, Kong Y, Lee S, Maye P, Yang SY (1996) Nucleotide sequences of MHC class I introns 1, 2, and 3 in humans and intron 2 in nonhuman primates. *Tissue Antigens* 47:498–511
- Cereb N, Maye P, Lee S, Kong Y, Yang SY (1995) Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens* 45:1–11
- Corbet GB (1978) The mammals of the palearctic region: a taxonomic review. *Publ Br Mus Nat Hist* 788:1–314
- Cresswell P (2005) Antigen processing and presentation. *Immunol Rev* 207:5–7
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482–489
- Del Cerro I, Marmi J, Ferrando A, Chashchin P, Taberlet P, Bosch M (2010) Nuclear and mitochondrial phylogenies provide evidence for four species of Eurasian badgers (Carnivora). *Zool Scr* 39:415–425
- Delpont W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457
- Figueroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature* 335:265–267
- Gessner C, Nakagawa S, Zavodna M, Gemmill NJ (2017) Sexual selection for genetic compatibility: the role of the major histocompatibility complex on cryptic female choice in Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity* 118:442–452
- Groot NGde, Otting N, Doxiadis GGM, Balla-Jhaghihoorsingh SS, Heeney JL, van Rood JJ et al. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc Natl Acad Sci USA* 99:11748–11753
- Gu X, Nei M (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol Biol Evol* 16:147–156
- Hancox M (1980) Parasites and infectious diseases of the Eurasian badger (*Meles meles* L.): a review. *Mamm Rev* 10:151–162
- Harasawa R, Orusa R, Giangaspero M (2014) Molecular evidence for hemotropic Mycoplasma infection in a Japanese badger (*Meles meles anakuma*) and a raccoon dog (*Nyctereutes procyonoides viverrinus*). *J Wildl Dis* 50:412–415
- Hornok S, Trauttwein K, Takács N, Hodžić A, Duscher GG, Kontschán J (2017) Molecular analysis of *Ixodes rugicollis*, *Candidatus* Neoehrlichia sp. (FU98) and a novel Babesia genotype from a European badger (*Meles meles*). *Ticks Tick Borne Dis* 8:41–44
- Huang J-F, Li L, Conran JG, Li J (2016) Phylogenetic utility of LEAFY gene in *Cinnamomum* (Lauraceae): gene duplication and polymerase chain reaction-mediated recombination. *J Syst Evol* 54:238–249
- Hughes AL (1995) Origin and evolution of HLA class I pseudogenes. *Mol Biol Evol* 12:247–258
- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170
- Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol Biol Evol* 6:559–579



- Hughes AL, Nei M (1992) Maintenance of MHC polymorphism. *Nature* 355:402–403
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* 32:415–435
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic Press, New York, pp 21–132
- Kaneko Y, Maruyama N, Macdonald DW (2006) Food habits and habitat selection of suburban badgers (*Meles meles*) in Japan. *J Zool* 270:78–89
- Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules—different yet the same. *Immunology* 6:411–424
- Kennedy LJ, Altet L, Angles JM, Barnes A, Carter SD, Francino O et al. (2000) Nomenclature for factors of the dog major histocompatibility system (DLA), 1998: first report of the ISAG DLA Nomenclature Committee. *Anim Genet* 31:52–61
- 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 et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* 31:217–219
- Klein J, Sato A, Nagl S, O’hUigín C (1998) Molecular trans-species polymorphism. *Annu Rev Immunol* 29:1–21
- Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annu Rev Genet* 41:281–304
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22:3096–3098
- Kurose N, Kaneko Y, Abramov VA, Siriaronrat B, Masuda R (2001) Low genetic diversity in Japanese populations of the Eurasian badger *Meles meles* (Mustelidae, Carnivora) revealed by mitochondrial cytochrome b gene sequences. *Zool Sci* 18:1145–1151
- Kydd JH, Case R, Winton C, MacRae S, Sharp E, Ricketts SL et al. (2016) Polarisation of equine pregnancy outcome associated with a maternal MHC class I allele: preliminary evidence. *Vet Microbiol* 188:34–40
- Lahr DJG, Katz LA (2009) Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* 47:857–866
- Li F, Luo ZH, Li CL, Li C, Jiang Z (2013) Biogeographical patterns of the diet of Palearctic badger: is badger an earthworm specialist predator? *Chin Sci Bul* 58:2255–2261
- Librado P, Rozas J (2009) DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Liu G, Zhang H, Sun G, Zhao C, Shang S, Gao X et al. (2017) Characterization of the peripheral blood transcriptome and adaptive evolution of the MHC I and TLR gene families in the wolf (*Canis lupus*). *BMC Genom* 18:584. <https://doi.org/10.1186/s12864-017-3983-0>
- Maibach V, Hans JB, Hvilsom C, Marques-Bonet T (2017) MHC class I diversity in chimpanzees and bonobos. *Immunogenetics* 69:661–676
- Marmesat E, Schmidt K, Saveljev AP, Seryodkin IV, Godoy JA (2017) Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the *Lynx* genus. *BMC Evol Biol* 17:158
- Marmi J, López-Giráldez F, Macdonald DW, Calafell F, Zholnerovskaya E, Domingo-Roura X (2006) Mitochondrial DNA reveals a strong phylogeographic structure in the badger across Eurasia. *Mol Ecol* 15:1007–1020
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:1–5
- Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrov* 21:98–102
- McAdam SN, Boyson JE, Liu XM, Garber TL, Hughes AL, Bontrop RE et al. (1994) A uniquely high-level of recombination at the hla-b locus. *Proc Natl Acad Sci USA* 91:5893–5897
- Miska KB, Harrison GA, Hellman L, Miller RD (2002) The major histocompatibility complex in monotremes: an analysis of the evolution of Mhc class I genes across all three mammalian subclasses. *Immunogenetics* 54:381–393
- Moreno B, Bolea R, Morales M, Martín-Burriel I, González CH, Badiola JJ (2015) Isolation and phylogenetic characterization of *Streptococcus halichoeri* from a European badger (*Meles meles*) with pyogranulomatous pleuropneumonia. *J Comp Pathol* 152:269–273
- Myers S, Freeman C, Auton A, Donnelly P, McVean G (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* 40:1124–1129
- 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, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152
- Nishita Y, Abramov AV, Kosintsev PA, Lin LK, Watanabe S, Yamazaki K et al. (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
- Nishita Y, Kosintsev PA, Haukialmi V, Väinölä R, Raichev EG, Murakami T et al. (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
- Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218–225
- Pan H-J, Wan Q-H, Fang S-G (2008) Molecular characterization of major histocompatibility complex class I genes from the giant panda (*Ailuropoda melanoleuca*). *Immunogenetics* 60:185–193
- Pokorny I, Sharma R, Tiedemann R (2010) MHC class I and MHC class II DRB gene variability in wild and captive Bengal tigers (*Panthera tigris tigris*). *Immunogenetics* 62:667–679
- Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol* 20:601–609
- Ross P, Buntzman AS, Vincent BG, Grover EN, Gojanovich GS, Collins EJ et al. (2012) Allelic diversity at the DLA-88 locus in golden retriever and boxer breeds is limited. *Tissue Antigens* 80:175–183
- Richman A, Herrera L, Nash D, Schierup M (2003) Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. *Genet Res* 82:89–99
- Saka T, Nishita Y, Masuda R (2017) Low genetic variation in the MHC class II DRB gene and MHC-linked microsatellites in endangered island populations of the leopard cat (*Prionailurus bengalensis*) in Japan. *Immunogenetics*. <https://doi.org/10.1007/s00251-017-1020-8>



- Sato JJ (2016) The systematics and taxonomy of the world's badger species—a review. In: Proulx G, Do Linh San E (eds) *Badgers: systematics, biology, conservation and research techniques*. Alpha Wildlife Publications, Sherwood Park
- Schaschl H, Suchentrunk F, Hammer S, Goodman SJ (2005) Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.). *Immunogenetics* 57:108–115
- Shawar SM, Vyas JM, Rodgers JR, Rich RR (1994) Antigen presentation by major histocompatibility complex class I-B molecules. *Annu Rev Immunol* 12:839–880
- Sin YW, Annavi G, Dugdale HL, Newman C (2014) Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Mol Ecol* 23:5072–5088
- Sin YW, Annavi G, Newman C, Buesching C, Burke T, Macdonald DW et al. (2015) MHC class II-assortative mate choice in European badgers (*Meles meles*). *Mol Ecol* 24:3138–3150
- Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T (2012a) MHC class II genes in the European badger (*Meles meles*): characterization, patterns of variation, and transcription analysis. *Immunogenetics* 64:313–327
- Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T (2012b) Evolution of MHC class I genes in the European badger (*Meles meles*). *Ecol Evol* 2:1644–1662
- Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34:126–129
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* 2:16
- Swindell SR, Plasterer TN (1997) SEQMAN. *Contig assembly*. *Methods Mol Biol* 70:75–89
- Takahashi K, Rooney AP, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. *J Hered* 91:198–204
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Tashima S, Kaneko Y, Anezaki T, Baba M, Yachimori S, Abramov AV, Saveljev AP, Masuda R (2011) Phylogeographic sympatry and isolation of the Eurasian badgers (*Meles*, Mustelidae, Carnivora): implications for an alternative analysis using maternally as well as paternally inherited genes. *Zool Sci* 28:293–303
- Venkataraman GM, Kennedy LJ, Little MTE, Graves SS, Harkey MA, Torok-Storb BJ et al. (2017) Thirteen novel canine dog leukocyte antigen-88 alleles identified by sequence-based typing. *HLA* 90:165–170
- Wagner JL (2003) Molecular organization of the canine major histocompatibility complex. *J Hered* 94:23–26
- Wagner JL, Sarmiento UM, Storb R (2002) Cellular, serological, and molecular polymorphism of the class I and class II loci of the canine major histocompatibility complex. *Tissue Antigens* 59:205–210
- Xiao J, Xiang W, Chai Y, Haywood J, Qi J, Qi J et al. (2016) Diversified anchoring features the peptide presentation of DLA-88\*50801: first structural insight into domestic dog MHC class I. *J Immunol* 197:2306–2315
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118
- York IA, Rock KL (1996) Antigen processing and presentation by the class I major histocompatibility complex. *Annu Rev Immunol* 14:369–396
- Yuhki N, Mullikin JC, Beck T, Stephens R, O'Brien SJ (2008) Sequences, annotation and single nucleotide polymorphism of the major histocompatibility complex in the domestic cat. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0002674>
- Zelano B, Edwards SV (2002) An MHC component to kin recognition and mate choice in birds: predictions, progress, and prospects. *Am Nat* 160:S225–S237
- Zhao M, Wang Y, Shen H, Li C, Chen C, Luo Z et al. (2013) Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. *BMC Evol Biol* 13:113. <https://doi.org/10.1186/1471-2148-13-113>
- Zhu Y, Sun DD, Ge YF, Yu B, Chen YY, Wan QH (2013) Isolation and characterization of class I MHC genes in the giant panda (*Ailuropoda melanoleuca*). *Chin Sci Bull* 58:2140–2147