Zoological Studies **62:**12 (2023)
doi:10.6620/ZS.2023.62-12

Zoological Studies

Open Access

Low Diversity of Major Histocompatibility Complex (MHC) Genes in Endangered Malayan Tapir (*Tapirus indicus*)

Nurul Adilah Ismail[1], Christina Seok Yien Yong[1], Simon Yung Wa Sin[2], and Geetha Annavi[1,*]

[1]*Department of Biology, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia.*
E-mail: adilahismail92@gmail.com (Ismail); chrisyong@upm.edu.my (Yong)
[2]*School of Biological Sciences, The University of Hong Kong, Pok Fu Lam Road, Hong Kong SAR.*
**Correspondence: E-mail: geetha@upm.edu.my (Annavi). Tel: +60397696621.*
E-mail: yungwa.sin@gmail.com (Sin)

Received 10 March 2022 / Accepted 3 January 2023 / Published 31 March 2023
Communicated by Jen-Pan Huang

The Malayan tapir (*Tapirus indicus*) is listed as Endangered on the IUCN Red List due to multiple threats such as habitat loss and human disturbance that have led to its population decline. This decline increases the risk of inbreeding, which could result in the reduction of genome-wide genetic variation and negatively affect the gene responsible for immune response *i.e.*, MHC gene. Class I and II MHC genes are responsible for encoding MHC molecules in the cells that recognise pathogenic peptides and present them to T-Cells on the cell surface for adaptive immune response. However, at present there is no study related to the MHC gene in Malayan tapir yet. This study characterises the MHC class I and II genes from seven individuals, investigates evidence of balancing selection and their relationships with homologous genes of other species. We identified at least one class I gene and four class II genes. Five sequences of alpha1 (α1) and four of alpha2 (α2) domains of class I alleles, two DRA, two DQA, three DRB and three DQB of class II alleles were isolated. α1 and α2 domains of class I and DRB domain of class II displayed evidence of selection with a higher rate of non-synonymous over synonymous substitutions. Within the DRB gene, 24 codons were found to be under selection where 10 are part of the codons forming the Antigen Binding Site. Genes sequences show species-specific monophyletic group formation except for class I and DRB genes with intersperse relationship in their phylogenetic trees which may indicate occurrence of trans-species polymorphism of allelic lineage. More studies using RNA samples are needed to identify the gene's level of expression.

Key words: MHC gene, Malayan tapir, Endangered mammals, MHC diversity, peptide-binding region.

BACKGROUND

The Major Histocompatibility Complex (MHC) gene is a multigene family in vertebrates that plays an important role in the adaptive immune system (Klein 1986; Janeway et al. 2001). This gene plays an essential role in encoding cell surface glycoprotein known as MHC molecules. The MHC molecules are encoded by two major classes of the MHC gene which are class I and II genes. In humans, class I genes are known as HLA-A, HLA-B, and HLA-C while the class II genes are DR, DP, DQ (Penn 2002; Blum et al. 2013). In mammals, when a cell is invaded with foreign pathogens such as viruses and bacteria, both class I and II MHC molecules will bind and present fragments of the foreign peptides on the cell's surface to T-cells and B-cells (Alberts et al. 2013). Once activated, these T cells and B cells will initiate an immediate immune response such as

Citation: Ismail NA, Yong CSY, Sin SYW, Annavi G. 2023. Low diversity of major histocompatibility complex (MHC) genes in endangered Malayan tapir (*Tapirus indicus*). Zool Stud **62:**12. doi:10.6620/ZS.2023.62-12.

lysis of the infected cells. High levels of polymorphism are common in the peptide-binding region (PBR) of the MHC molecules. The amino acid variation within the allele sequences encoded for MHC molecules affects the PBR binding specificity to pathogens. Thus, high variation of the genes allows for a wider range of pathogen recognition (Sommer 2005).

Perhaps due to its important function in recognising a wide range of pathogens, the MHC is the most polymorphic gene in vertebrates (Janeway et al. 2001). The high diversity in the gene is attributed to balancing selection, a type of pathogen-mediated selection that maintains the high allelic frequency and nucleotide diversity in the population (Hughes and Hughes 1995). Heterozygote advantage, rare allele advantage and fluctuating selection were hypothesised to be the driving forces behind the pathogen-mediated selection. There is also an established theoretical framework that supports the idea that MHC diversity is driven by any a or a combination of the three mechanisms (Spurgin and Richardson 2010). It is difficult to pinpoint exactly which mechanism drives selection in a particular species. However, it is possible to detect the historical selection that occurs on the gene (Bernatchez and Landry 2003). In most protein coding genes where selection is neutral, the rate of synonymous nucleotide substitutions (substitution that does not result in change in amino acid) is greater than the non-synonymous substitutions (substitutions that result in changes in amino acid). This is because non-synonymous substitutions tend to change the amino acid, therefore are likely to be deleterious (Graur and Li 2000). However, the MHC gene that encodes for the PBR of the MHC molecule has been observed to display a higher rate of non-synonymous substitutions than synonymous substitutions. The higher rates in this gene may signal that the allele that is under selection could be advantageous in the population (Li 1993). This scenario is common in the MHC gene that encode for different species across multiple taxonomy such as European Badger *Meles meles* (Sin et al. 2012a b), Koala *Phascolarctos cinereus* (Cheng et al. 2018) and Spotted pardalote *Pardalotus punctatus* (Balasubramaniam et al. 2017). Balancing selection could result in retention of a large number of alleles in populations for a long period of time, in which the allele is passed down even after species speciation and results in trans-species polymorphism (Klein 1986).

In this study, we characterise the PBR of class I and II genes in Malayan tapir *(Tapirus indicus*). Malayan tapir is one of the five tapir species that belongs to the family Tapiridae and Order Perissodactyla. Currently, this species is listed as an Endangered species on the International Union for Conservation of Nature (IUCN) Red List and in CITES Appendix I due to multiple factors predominated by habitat loss and human disturbance. Currently, the population of Malayan tapir in the wild is estimated to be only around 2000–2500 individuals, which calls for more efforts to conserve this mammal (Lynam et al. 2008; Traeholt et al. 2016). Existing conservation efforts for this species include captive breeding and wild population monitoring. However, one of the most challenging factors that make it harder to recover from its low population number is that Malayan tapir has a slow reproduction rate whereby they can generally produce one calf every two years after a long gestation period (390–395 days) (Barongi 1993). The small population number further increases the risk of inbreeding, which could lead to inbreeding depression in the population (Hughes and Hughes 1995; Benton et al. 2018). Inbreeding, which is characterised by the loss of genetic variability including at the MHC loci, could decrease their ability to fight pathogens and increase their susceptibility towards diseases (Hedrick and Miller 1994; Keller and Waller 2002; Spielman et al. 2004).

One such example of reduced MHC diversity due to inbreeding that led to negative impacts is in the Tasmanian devil. MHC Class I genes were then unable to recognise infectious tumors as foreign, allowing for aggressive invasion into naïve individuals that led to significant population declines. There is also reported high mortality in inbred cheetahs due to coronavirus-associated feline infectious peritonitis (O'Brien and Evermann 1988). Reduction in MHC variation, at least in the MHC class I gene, has also been observed in this species (Schwensow et al. 2019). In the case of Malayan tapir, there is very little information on the MHC gene, and our understanding on the population fitness in relation to the MHC gene is far from complete. Therefore, characterisation of the gene from this study would facilitate our understanding of the evolutionary process that influences MHC gene diversity and serve as a basis for further study on pathogen resistance in this species. Although this study lacks RNA information, the gDNA sequence remains essential in providing preliminary information of the MHC allele and its variation in Malayan tapir with a focus on the important part of the MHC molecule: the peptide-binding region.

Objectives

This study aims to: 1) characterise the MHC class I and II genes that encode for PBR in Malayan tapir and test for evidence of selection based on gDNA alleles; (2) perform phylogenetic analysis to investigate whether Malayan tapir MHC genes belong to a monophyletic group or if there is an occurrence of trans-species

polymorphism.

MATERIALS AND MATERIALS

Sample collection and DNA isolation

Whole blood samples were collected in 2016 from a total of seven healthy captive Malayan tapirs: one sample from Zoo Negara, Malaysia; the remaining six samples from Sungai Dusun Wildlife Reserve, Malaysia (Table S1). The individuals were assumed to be unrelated and coming from different populations by checking their birth and transfer histories to the respective captive centres. Approximately 3 ml of blood was taken by jugular venipuncture by the veterinarians from Department of Wildlife and National Park (DWNP) and deposited into blood collection tubes containing EDTA. Collected blood samples were allocated into several 1.5 ml tubes with approximately 1 ml per tube for storage to avoid multiple freeze-thaw cycles of the samples for DNA isolation. The samples were stored on ice temporarily while in the field and were then immediately transported to the lab upon sampling completion. All the samples were transferred into a -20°C refrigerator until genomic DNA (gDNA) was isolated. gDNA was isolated from approximately 500 µl of blood samples using the QIAamp® DNA Mini Kit (Qiagen, Germany) following the manufacturer's spin protocol. Extracted DNA samples were quantified using Quantus™ Fluorometer stained with ONE dsDNA dye (Promega, USA). All gDNA samples quantified were above 3.0 ng/µl.

Primers Design

To amplify the peptide binding region of the class I (exon 2 and 3) and class II (exon 2) MHC genes in Malayan tapir, both published and designed primers were tested. Published primers from horse (Albright-Fraser et al. 1996; Fraser and Bailey 1998; Hedrick and Miller 1994; Kurtz et al. 2010) were tested. Oligonucleotide primers were designed to recognise the highly conserved peptide binding region. The primers were designed using MEGA 5 (Tamura et al. 2011) and Primer 3 Plus (Untergasser et al. 2007) based on consensus alignment of the closely and distantly related species sequences from GenBank (Table 1a and 1b).

PCR Amplification

Each of the primers were tested using PCR amplification in a 20 µl single reaction containing 1x MyTaq Red Mix (Bioline, Germany), 3–15 ng of template DNA, and 0.5–0.8 µM of primer mix using a touchdown PCR profile. The PCR cycle started with an incubation period of 2 mins at 92°C, followed by 35 amplification cycles started with 92°C for 30s, annealing temperature decreased -1°C per cycle starting at 60°C–50°C for 1 min with the rest of the cycle kept constant at the last annealing temperature, extension at 72°C for 30s, and ended with a single cycle of final extension at 72°C for 5 mins.

To confirm success of PCR amplification, 3 µl of PCR products were visualised on 2% agarose gel stained with RedSafe Nucleic Acid Staining Solution. The PCR products that contained the band of expected sizes were purified using Wizard® PCR Clean-Up

Table 1. (a) GenBank accession numbers for sequences of closely and distantly related species of Malayan tapir; (b) Primers used for MHC Class I and II amplification. F = forward, R = reverse, T_a = annealing temperature, bp = base pair, td = touchdown

	Primer name	Primer Sequence	Product size (bp)	Ta (°C)	Region amplified	Source
Class I	F: Class1_F1	GTGGACGACACGCAGTTC	791	60–50 (td)	exon2-exon3	This study
	R: Class1_R1	GTGAACAAATCTCGCATC		60–50 (td)		This study
	F: Class1_F2	GGTCTCCCGGTTTCCAGGG	275	60–50 (td)	exon 3	This study
	R: Class1_R2	GCGCTGCAGCGTCTCC		60–50 (td)		
Class II	F: DRA_F2	TTCTATCTGAACCCTGACC	175	55	exon 2 DRA	This study
	R: DRA_R1	GTTGGCTTTGTCCACAGCTA		57		This study
	F: C2DRB_LA31	GATGGATCCTCTCTCTGCAGCACATTTCCT	308	60–55 (td)	exon 2 DRB	Hedrick et al (1999)
	R: C2BRB_LA32	CTTGAATTCGCGCTCACCTCGCC GCTG		60–55 (td)		Hedrick et al (1999)
	F: C2DQA_2E	CTGAICACITTGCCTCCTATG	247	55	exon 2 DQA	Fraser and Bailey (1998)
	R: C2DQA_2F	TGGTAGCAGCAGIAGIGTTG		53		Fraser and Bailey (1998)
	F: DQB_F2	TGCTACTTCACCAACGG	205	55	exon 2 DQB	This study
	R: DQB_R3	GTAGTTGTGTCTGCACAC		55		This study

System (Promega, USA) following manufacturer's protocol. Purified PCR products were checked again on 2% agarose gel before being ligated and cloned. Purified PCR products were ligated with pGEM®-T Easy Vector (Promega, USA) into a 10 µl ligation reaction with the following modified protocol: 5 µl 2X Rapid Ligation Buffer, 0.5 µl pGEM®-T Easy Vector, 4.0 µl purified PCR product, 0.5 µl T4 DNA ligase. The ligation reactions were left overnight in 4°C. 5 µl of ligated product was transformed into 20 µl of JM109 High-Efficiency Competent Cells (*Escherichia coli*) in 1.5 ml tube (to produce multiple copies of a recombinant DNA molecule) and 980 µl of SOC Medium were added bringing the mixture to approximately 1 ml following manufacturer's instructions. The bacteria with PCR insert were cultured in an incubator with shaking at 37°C for 1 hour and 30 minutes before plating onto LB agar containing ampicillin (50 ng/ml) and X-gal (40 mg/ml). The plates were then placed in an incubator at 37°C for 16 to 18 hours for the colonies to grow. Plates with grown colonies were then stored at 4°C overnight to allow for further formation of blue-white bacteria colonies.

Approximately 9–20 white colonies were randomly picked from each plate and colony PCR was performed on each colony using M13 primers (forward: 5'-d(GTTTTCCCAGTCACGAC)-3', reverse:5'-d(CAGGAAACAGCTATGAC)-3'). Colony PCR was performed in a 10 µl reaction containing 1X MyTaq™ Red Mix, 0.5 uM of M13 primer mix. The colony PCR cycle started with an incubation period of 2 mins at 92°C, followed by 35 amplification cycles with each starting with a denaturation temperature of 92°C for 30s, annealing temperature at 55°C for 30s, extension at 72°C for 30s. The PCR ended with a single cycle of final extension at 72°C for 5 mins. 2 µl of colony PCR products were visualised on 2% agarose gel and colonies with expected sizes were sequenced (forward, reversed or both directions) by MyTACG Bioscience Enterprise, Malaysia.

Chromatograms of the sequences obtained were analysed using Finch TV 1.4.0 (Geospiza, Inc., Seattle, WA, USA). Identical sequences in the chromatograms were derived from a minimum of two individuals or independent PCR reactions of the same individual were identified as true alleles. Single unique sequences that may indicate possible chimeras or PCR errors were excluded from further analysis. DNA sequences obtained from Malayan tapir in this study were assigned GenBank accession numbers: MK432928-MK432945 and MK482362. Sequences from NCBI BLAST (Altschul et al. 1990) were retrieved and compared with the obtained sequences in this study. The nucleotide sequences were edited in ClustalX (Thompson et al. 2003) and MEGA 5 (Tamura et al. 2011).

Data analysis

Selection analysis

Selection at the amino acid level for the MHC genes was measured as the rates of nonsynonymous (d_N) and synonymous (d_S) substitutions per codon site by using DnaSP 4.0 (Rozas et al. 2003) and MEGA 7 (Kumar et al. 2016). The rates were measured in accordance with the Nei and Gojobori method (Nei and Gojobori 1986) with Jukes and Cantor correction (Jukes and Cantor 1969). Standard errors were obtained by bootstrap procedure with 1000 replicates. Amino acids that made up the antigen-binding site (ABS) and non-antigen-binding-site (non-ABS) were identified based on Reche and Reinherz (2003). Codons that formed the ABS and non-ABS in the MHC genes were respectively calculated for synonymous and nonsynonymous rates. CODEML program within the PAML 4.4b software (Yang 2007) was used to identify the positive selection sites (PSS) in the class I α1 and α2 domains as well as class II α1 and β1 domains (both in DR and DQ). An Ω value, which is nonsynonymous over synonymous substitutions (dN /$\boldsymbol{d}$S), larger than 1 indicates PSS within the domains in class I and II genes.

Codon-based likelihood analysis was used to test for evidence of positive selection, using several models. Within the CODEML program in PAML, null models and alternative models of nucleotide substitutions were applied and compared. The null models consist of M0-one ratio, M1a-nearly neutral, and M7-beta. The null models consist of parameters that reflect neutral evolution of nucleotide substitution. The nested/alternative models, which are M3-discrete, M2a-positive selection, M8-beta and ω, consist of parameters that allow for positive selection. The parameters set for both null and alternative models are detailed in (Yang et al. 2005). The likelihood ratios (2ΔlnL) of null and alternative models were compared to an χ^2 distribution to determine whether the alternative model provided a significantly improved fit, versus the null model. CODEML was also used to calculate Bayes Empirical Bayes (BEB) posterior probabilities (Yang 2007) to identify codons under positive selections, for comparisons of null and alternative models (M1a versus M2a and M7 versus M8). Codons with BEB posterior probabilities greater than 0.95 indicate the positive selection.

Phylogenetic analysis

Malayan tapir class I (exon 2 and exon 3) and class II (exon 2 of DRA, DRB, DQA, and DQB) sequences were aligned with sequences from closely related species available in GenBank using BioEdit 7.1.3.0 with ClustalW algorithm (Hall 1999) for phylogenetic analysis. Domain borders for the Malayan tapir MHC gene sequences were assigned based on the homologous sequences to its respective HLA genes available on IMGT/HLA database (Robinson et al. 2014). Bayesian phylogenetic inference was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). For each dataset, a Markov chain Monte Carlo (MCMC) search was initiated for at least 1,000,000 generations and sampling for every 100 generations and the first 25% were discarded as "burn in". The standard deviation of split frequencies converged to a value of less than 0.05. Two separate analyses and four independent chains were executed for all datasets. Evidence of convergence was also checked by plotting the likelihood scores against generations.

The gene sequence alignments were run into FindModel (https://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) to find the best fit model for nucleotide substitution. Models were selected based on the lowest Akaike information criterion (AIC) value to better fit the data (Akaike 1974).

RESULTS

Gene Characterisation

Five exon 2 and four exon 3 of class I MHC gene sequences, alongside two DRA, three DRB, two DQA and three DQB sequences of class II MHC were successfully isolated from gDNA samples of seven Malayan tapir individuals using the primers detailed in table 1. More than two class I alleles were detected in T3 (exon 2) and T6 individuals (exon 2 and 3) (Table 2), whereas more than one class I alleles were observed in T3 and T6 individuals (Table 3), indicating the possibility of the presence of multiple class I MHC loci in Malayan tapir. Not more than two alleles were observed within each individual for all class II MHC genes which could indicate a single locus of those genes. A comparison of sequences deduced from the seven Malayan tapirs show high amino acid polymorphism was observed in the exon 2 and exon 3 of class I MHC alleles (Table 4). Meanwhile, in class II MHC alleles, the DRA gene showed the lowest polymorphism with only one nucleotide difference observed. The highest polymorphism was observed in the DRB gene with 24 polymorphic amino acids. Both class I and class II genes in Malayan tapir displayed higher non-synonymous amino acid substitutions than synonymous amino acid substitutions in the compared sequences.

Table 2. Presence of MHC class I allele in Malayan tapir individuals. The individuals are denoted as T1-T7

Alleles /	Exon 2				
Individuals	Tain_C1_01	Tain_C1_02	Tain_C1_03	Tain_C1_04	Tain_C1_05
T 1	x				
T 2	x				
T 3		x		x	x
T 4	x				
T 5	x				x
T 6	x		x		x
T 7	x	x			

Alleles /	Exon 3			
Individuals	Tain_C1_06	Tain_C1_07	Tain_C1_08	Tain_C1_09
T 1				x
T 2	x			x
T 3	x			x
T 4				x
T 5			x	x
T 6	x	x	x	x
T 7				x

Based on the aligned sequences with closely related species, Malayan tapir class I MHC allele sequences Tain_C1_01 and Tain_C1_02 (exon 2, Fig. S1), and Tain_C1_06 and Tain_C1_07 (exon 3, Fig. S2) showed amino acid differences that were distinct from orthologous sequences, while the remaining Malayan tapir class I sequences showed high amino acid similarity to the closely related species. In Malayan tapir class II gene sequences, the two sequences of DRA gene (Tain_DRA01 and Tain_DRA02) showed a high similarity between themselves and also with orthologous genes from closely related sequences (Fig. S3). High similarity was also observed between DQA sequences (Tain_DQA01 and Tain_DQA02). However, orthologous genes from closely related species showed lower similarity in the DQA gene (Fig. S4). Malayan tapir DRB gene also showed moderate amino acid similarity to its closely related species (Fig. S5). Malayan tapir DQB gene showed moderate amino acid similarity to its closely related species (Fig. S6).

Table 3. Presence of MHC class II allele in Malayan tapir individuals. The individuals are denoted as T1-T7

Alleles / Individuals	Class II				
	DRA		DRB		
	Tain_DRA01	Tain_DRA02	Tain_DRB01	Tain_DRB02	Tain_DRB03
T1	x				x
T2	x		x	x	
T3	x			x	x
T4		x		x	
T5	x	x			x
T6	x		x	x	
T7		x	x	x	

Alleles / Individuals	Class II				
	DQA		DQB		
	Tain_DQA01	Tain_DQA02	Tain_DQB01	Tain_DQB01	Tain_DQB02
T1		x	x		
T2		x		x	
T3	x	x	x	x	
T4		x			
T5	x	x	x		
T6		x			x
T7		x			x

Table 4. Sequence polymorphism of MHC Class I and II genes

	Class I		Class II			
Exon	Exon 2	Exon 3	Exon 2			
Domain	α1	α2	DRα	DQα	DRβ	DQβ
Sequence for comparison	5	4	9	2	3	3
Variable sites	94	85	7	1	35	8
Mutations	104	94	7	1	40	8
Synonymous	19	23	3	0	4	3
Nonsynonymous	43	54	3	1	36	5
No. of amino acid	91	95	63	82	103	69
Polymorphic amino acid residue	46	42	3	1	24	5

Selection Analysis

Within Malayan tapir class I sequences, the rate of non-synonymous (d_N) to synonymous substitutions (d_S) was observed to be higher in codon coding for ABS compared to non-ABS in both exons (Table 5). Exon 2 and exon 3 recorded ω values greater than 1 at 1.37 and 1.18 respectively. In class II genes, only the DRB gene shows a higher rate of non-synonymous substitutions (ω = 3.52) indicating the positive selection in this domain (Table 5). Further analysis using PAML inferred PSS only in DRB class II gene and none in class I genes (Table 6). Comparisons between the M1a and M2a models and the M7 and M8 models show significant results when compared to χ^2 distribution inferred PSS of 24 amino acids whereby ten are identified as codons forming the ABS within the DRB sequence (Table 7, see Fig. 5).

Phylogenetic Analysis

Based on the AIC value (Table S2), class I genes were analysed with a general time-reversible plus gamma model (GTR+Γ) for the exon 2 dataset and Hasegawa-Kishino-Yano plus gamma model (HKY+Γ) for exon 3. For class I genes, a Markov chain Monte Carlo (MCMC) search was initiated with random trees and ran for 3,000,000 generations, with a sampling of every 100 generations. Class II genes were analysed with a general time-reversible plus gamma model (GTR+Γ) for DRB, DQB and beta domain datasets, while Kimura two parameters (K80) were applied to DRA, DQA and alpha domain datasets. For class II genes, a MCMC search was initiated with random trees and ran for 1,000,000 generations with a sampling frequency of every 100 generations.

In class I genes, the phylogenetic trees of alpha 1 (five sequences) and alpha 2 (four sequences) domains highlight that the sequences from *Tapirus indicus* formed two sub clades (Fig. 1). The Tain_C1_01 and Tain_C1_02 formed closer clades to the homologous sequences of horses and humans. The other two sequences (Tain_C1_03 and Tain_C1_04) form monophyletic clades closer to rhinoceros and bovine species. The clustering may indicate different loci of the alpha1 region, while the alpha 2 region of *Tapirus indicus*, excluding the Tain_C1_08 sequence, formed a monophyletic group (Fig. 2).

Table 5. Rates of non-synonymous (dN) and synonymous (ds) substitutions for antigen-binding site (ABS) and non-ABS, and combined (ABS and non-ABS) at the Malayan tapir *Tapirus indicus* MHC class I loci as determined in PAML. ω indicate ratio of non-synonymous to synonymous nucleotide substitutions

Region	Position	Number of codons	d_N	d_S	ω
	α1 (exon2)				
	ABS	12	0.45 ± 0.13	0.33 ± 0.18	1.37
	Non-ABS	68	0.38 ± 0.06	0.42 ± 0.08	0.89
Class I	Combined	80	0.39 ± 0.05	0.41 ± 0.08	0.96
	α2 (exon3)				
	ABS	9	0.42 ± 0.17	0.36 ± 0.36	1.18
	Non-ABS	74	0.35 ± 0.05	0.42 ± 0.08	0.82
	Combined	83	0.35 ± 0.04	0.42 ± 0.08	0.85
	DRB				
	ABS	16	0.48 ± 0.10	0.14 ± 0.09	3.52
	Non-ABS	86	0.10 ± 0.03	0.04 ± 0.03	2.81
	Combined	102	0.16 ± 0.03	0.05 ± 0.03	3.10
	DQB				
	ABS	13	0.04 ± 0.04	0.00 ± 0.00	0.00
	Non-ABS	56	0.03 ± 0.02	0.13 ± 0.10	0.27
Class II	Combined	69	0.04 ± 0.02	0.11 ± 0.08	0.34
	DQA				
	ABS	13	0.00 ± 0.00	0.00 ± 0.00	0.00
	Non-ABS	69	0.01 ± 0.01	0.00 ± 0.00	0.00
	Combined	82	0.01 ± 0.01	0.00 ± 0.00	0.00
	DRA				
	ABS	15	0.00 ± 0.00	0.00 ± 0.00	0.00
	Non-ABS	40	0.01 ± 0.01	0.07 ± 0.07	0.18
	Combined	55	0.00 ± 0.00	0.05 ± 0.05	0.00

Table 6. Null (M1a and M7) and alternative (M2a and M8) models comparison with their parameters determined in PAML for class I and II DRB gene in Malayan tapir

Domain	Model	lnL	Parameter estimates(s)	LRT	2ΔlnL	*P* value
Class I α1 (exon2)	M1a	-722.30	$p_0 = 0.60\ \omega_0 = 0.10\ p_1 = 0.40\ \omega_1 = 1.00$	M1a vs M2a	0.06	< 0.5 (NS)
	M2a	-722.27	$p_0 = 0.62\ \omega_0 = 0.11\ p_1 = 0.32\ \omega_1 = 1.00\ p_2 = 0.06\ \omega_2 = 1.87$			
	M7	-722.53	$p = 0.31\ q = 0.44$	M7 vs M8	0.66	< 0.5 (NS)
	M8	-722.20	$p_0 = 0.84\ p = 0.60\ q = 1.67\ p_1 = 0.16\ \omega = 1.80$			
Class I α2 (exon3)	M1a	-736.27	$p_0 = 0.25\ \omega_0 = 0.00\ p_1 = 0.75\ \omega_1 = 1.00$	M1a vs M2a	1.44	> 0.9 (NS)
	M2a	-735.56	$p_0 = 0.31\ \omega_0 = 0.00\ p_1 = 0.00\ \omega_1 = 1.00\ p_2 = 0.69\ \omega_2 = 1.45$			
	M7	-736.43	$p = 0.03\ q = 0.01$	M7 vs M8	1.74	> 0.9 (NS)
	M8	-735.56	$p_0 = 0.31\ p = 0.01\ q = 2.29\ p_1 = 0.69\ \omega = 1.45$			
Class II DRB (exon 2)	M1a	-582.01	$p_0 = 0.50\ \omega_0 = 0.00\ p_1 = 0.50\ \omega_1 = 1.00$	M1a vs M2a	19.56	< 0.001*
	M2a	-572.23	$p_0 = 0.88\ \omega_0 = 0.78\ p_1 = 0.00\ \omega_1 = 1.00\ p_2 = 0.12\ \omega_2 = 21.61$			
	M7	-585.66	$p = 55.96\ q = 0.01$	M7 vs M8	26.86	< 0.001*
	M8	-572.23	$p_0 = 0.88\ p = 99\ q = 27.15\ p_1 = 0.12\ \omega = 21.62$			

Table 7. Positive selected sites were identified in models M2a and M8 by Bayes Empirical Bayes (BEB) Posterior Probabilities. Asterisk mark indicate codon forming ABS

Model LRT	Codon number	Amino acid	Probability ω > 1	Mean ω ± SE
M1a versus M2a	16	H	0.99	9.19 ± 1.53
	31	D	0.99	9.20 ± 1.50
	33	Y	0.99	9.18 ± 1.56
	74	K	0.99	9.18 ± 1.15
	89	G	0.96	8.89 ± 2.16
M7 versus M8	12	V	0.96	1.47 ± 0.16
	13	Q	0.96	1.48 ± 0.16
	14	V	0.96	1.50 ± 0.00
	16	H	0.98	1.49 ± 0.13
	28	R	0.96	1.51 ± 0.00
	29	F*	0.96	1.50 ± 0.00
	31	D*	0.98	1.49 ± 0.13
	33	Y*	0.98	1.49 ± 0.13
	34	F	0.96	1.50 ± 0.00
	37	R	0.96	1.51 ± 0.00
	40	Y*	0.96	1.50 ± 0.00
	41	V*	0.96	1.51 ± 0.00
	50	Y*	0.96	1.50 ± 0.00
	52	P	0.96	1.51 ± 0.00
	60	D*	0.96	1.51 ± 0.00
	64	W*	0.96	1.51 ± 0.00
	70	L*	0.96	1.51 ± 0.00
	73	Q	0.96	1.47 ± 0.16
	74	K	0.98	1.49 ± 0.13
	81	Y*	0.96	1.50 ± 0.00
	89	G*	0.96	1.48 ± 0.16
	90	E	0.96	1.51 ± 0.00
	91	S	0.96	1.51 ± 0.00
	93	T*	0.96	1.51 ± 0.00

In class II genes, the two DRA sequences from *Tapirus indicus* formed a monophyletic group with Tapirus species (Fig. 3). The sequences also formed groups with sequences from rhinoceros species and were interspersed with the human DRA sequence. These sequences formed a distinct group from the horses' sequences, which formed a monophyletic group. Meanwhile, the two *Tapirus indicus* DQA sequences formed a monophyletic group with equine sequences (Fig. 4). In beta domain genes, *Tapirus indicus* DRB gene sequences showed interspersed clustering with sequences from bovine (Fig. 5) while *Tapirus indicus* DQB sequences formed monophyletic groups (Fig. 6).

DISCUSSION

Gene Characterisation

In this study we were able to isolate almost full-length exon 2 and exon 3 of class I genes and a partial majority exon 2 of class II genes (DRA, DRB, DQA and DQB). Our inability to amplify the rest of the gene sequences was due to the fact that these sequences would not amplify as readily with the primers that we designed or the ones that were published. Our designed primers were generated from the limited conserved regions of closely related species (mainly horses and rhinoceros) to increase the probability of working primers for amplification. Therefore, this limits the ability to amplify longer gene sequence lengths and increases the probability of missing more alleles. Sequencing data from the amplification using these primers also produces many unique single sequences from the class I and II genes. As we could not reproduce the sequences with independent PCR, these unique sequences were not included in further analysis in this study.

The isolated sequences of the Malayan tapir class I exon 2 and exon 3 regions display conserved characteristics that are consistent with common antigen recognition sites in many species that could represent

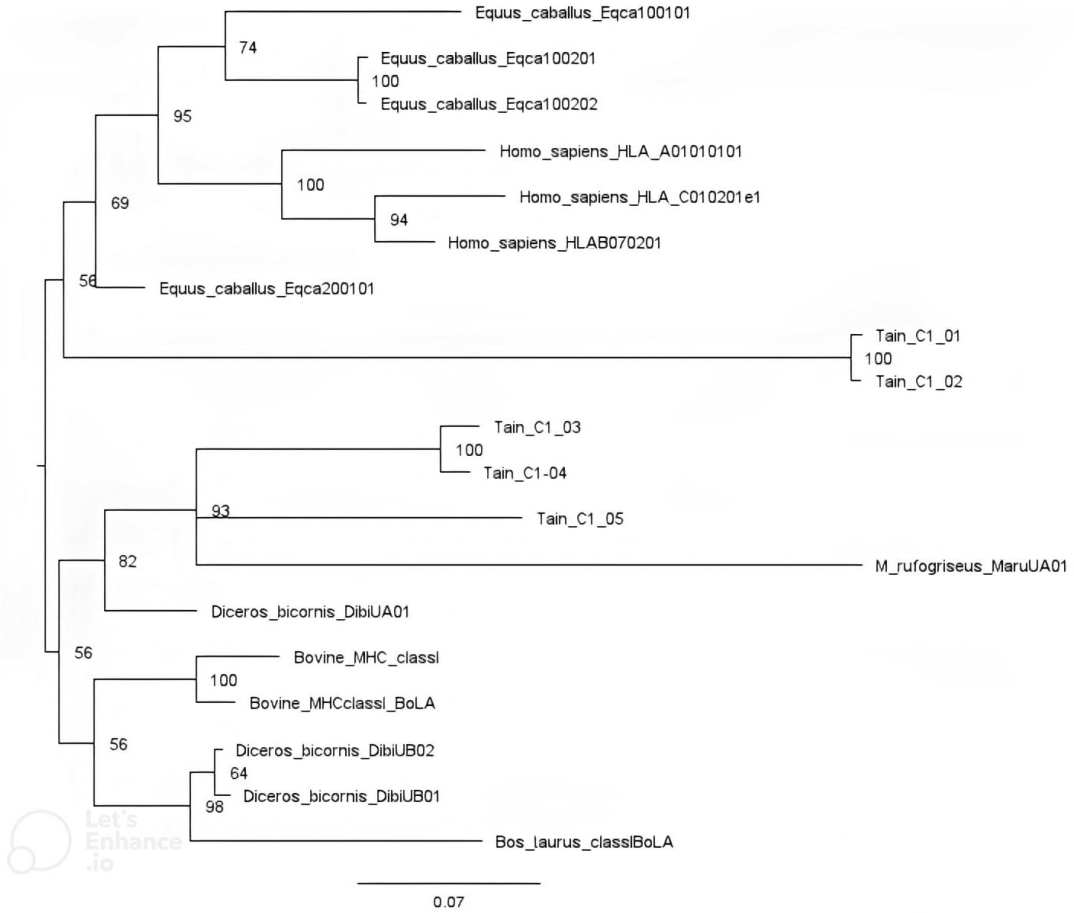

Fig. 1. Phylogenetic tree of MHC class I α1 domain (exon2) sequences from Malayan tapir *Tapirus indicus*, equids, rhinoceros and other mammals including human and bovine. Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences in this tree are Tain_C1_01, Tain_C1_02, Tain_C1_03, Tain_C1_04, and Tain_C1_05.

functional molecules (Ellis et al. 1995). An example would be the presence of N-linked glycosylation at position 88 in exon 2 and conserved cysteine residues at position 11 and 78 in exon 3. The conserved cysteine residues within exon 3 function as sites for intra domain disulphide bond formation in peptide binding (Ellis et al. 1995). However, although the sequences contain conserved characteristics of functional molecules, the degree of functionality of these alleles could not be proved. Apart from the absence of cDNA data, there are presences of stop codons and indels within the aligned Malayan tapir sequences with its closely related species. For example, within exon 2, indels can be observed at amino acid position 51, 52, and 53 (Tain_C1_01 and Tain_C1_02) of exon 2 sequence. The exon 3 sequences also show similar occurrences of indels in multiple positions such as at 21, 22 and 23. The presence of indels in these positions may indicate the sequences are pseudogene alleles. However, classification of loci below was done simply by assuming that more similar alleles originated from the same locus and lack functionality confirmation. More alleles could have been missed as the primers had captured alleles with the presence of indels. Therefore, more studies are required to assess the functionality and allele assignment of these sequences to their class I loci.

Malayan tapir class II genes contain no obvious stop codon or in-frame mutation. However, we could not rule out the possibility of indels or stop codons in the remainder of the class II gene sequences. These sequences contain all conserved residues expected of a functional peptide binding region of class II genes as identified in Reche and Reinherz (2003). This can be observed in the conserved alanine and valine residues in the α-genes (DRA position 52 and 53, DQA position 63 and 64) and also in conserved threonine residue in β-genes (DRB position 79 and DQB position 63). The DRB region shows the highest polymorphism while

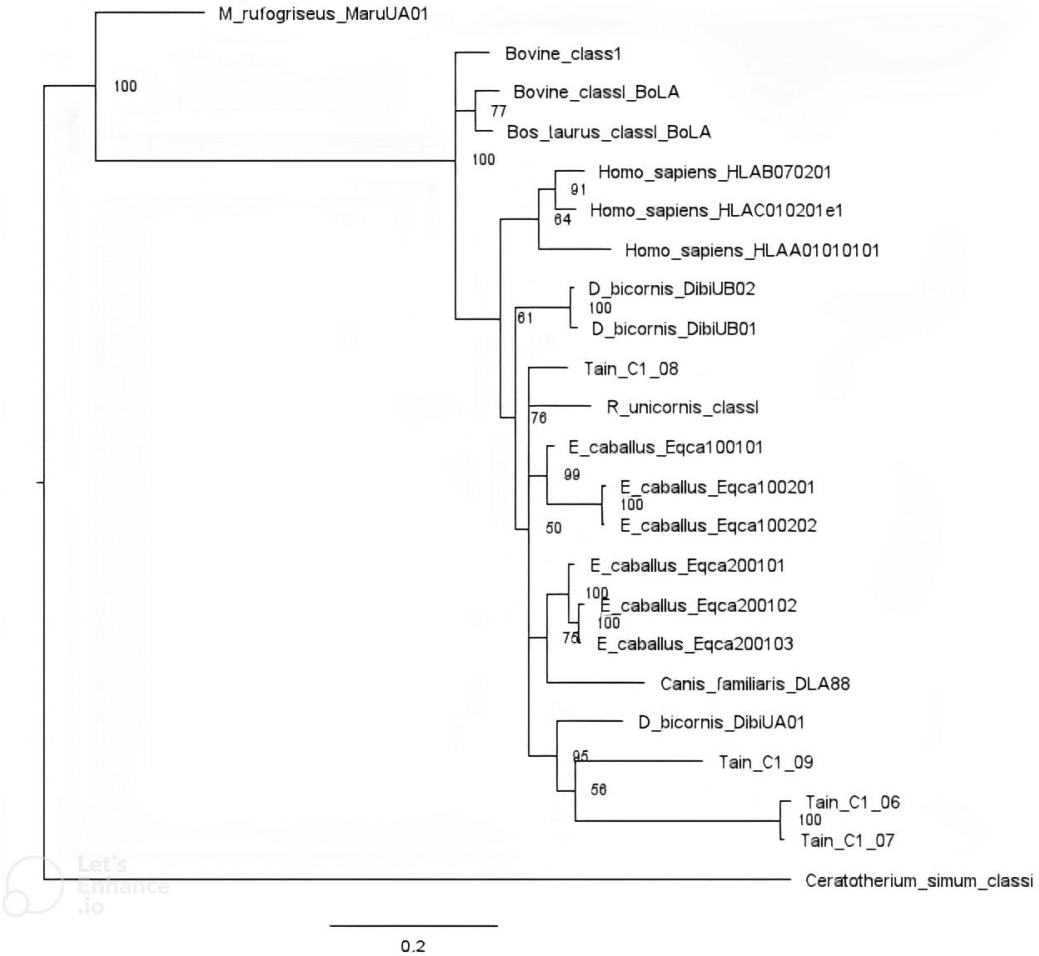

Fig. 2. Phylogenetic tree of MHC class I α2 domain (exon3) sequences from Malayan tapir *Tapirus indicus*, equids, rhinoceros and other mammals including human and bovine. Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences in this tree are Tain_C1_06, Tain_C1_07, Tain_C1_08, and Tain_C1_09.

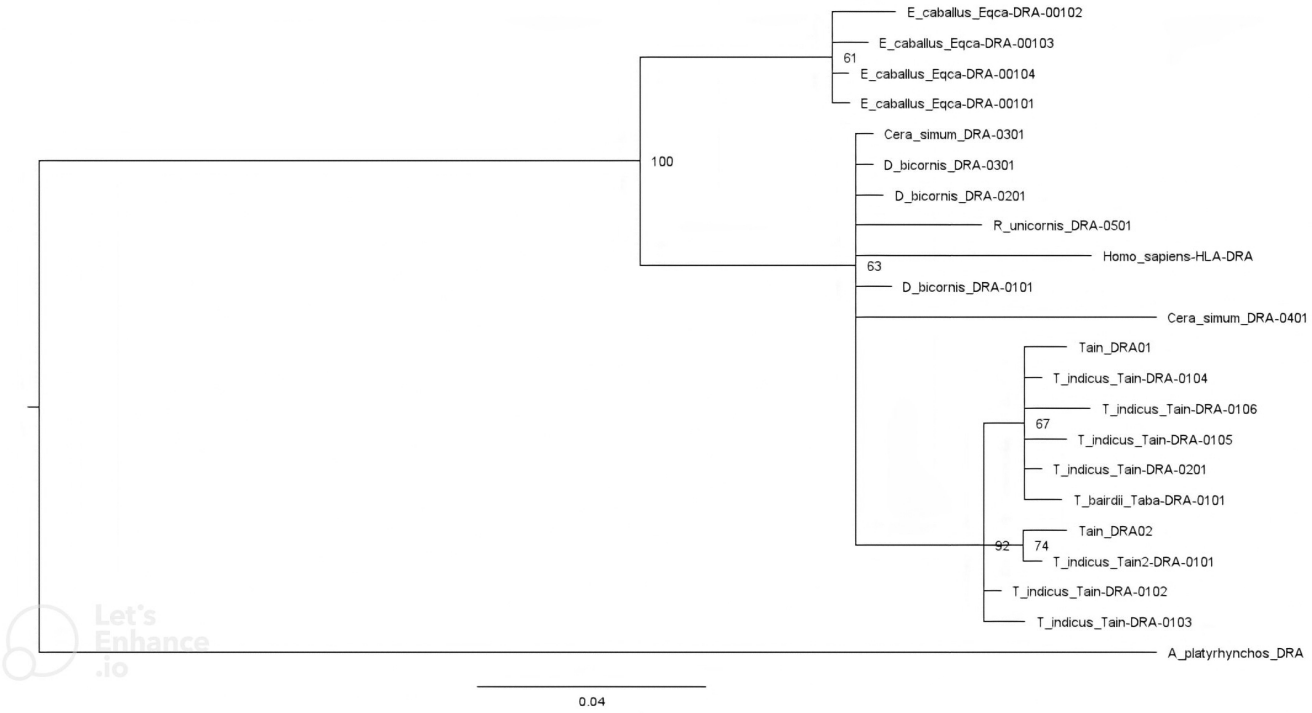

Fig. 3. Phylogenetic tree of MHC class II DRα domain (exon2) sequences from Malayan tapir *Tapirus indicus*, equids, rhinoceros and other mammals including human and fish (outgroup). Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences from this study in this tree are Tain_DRA01 and Tain_DRA02.

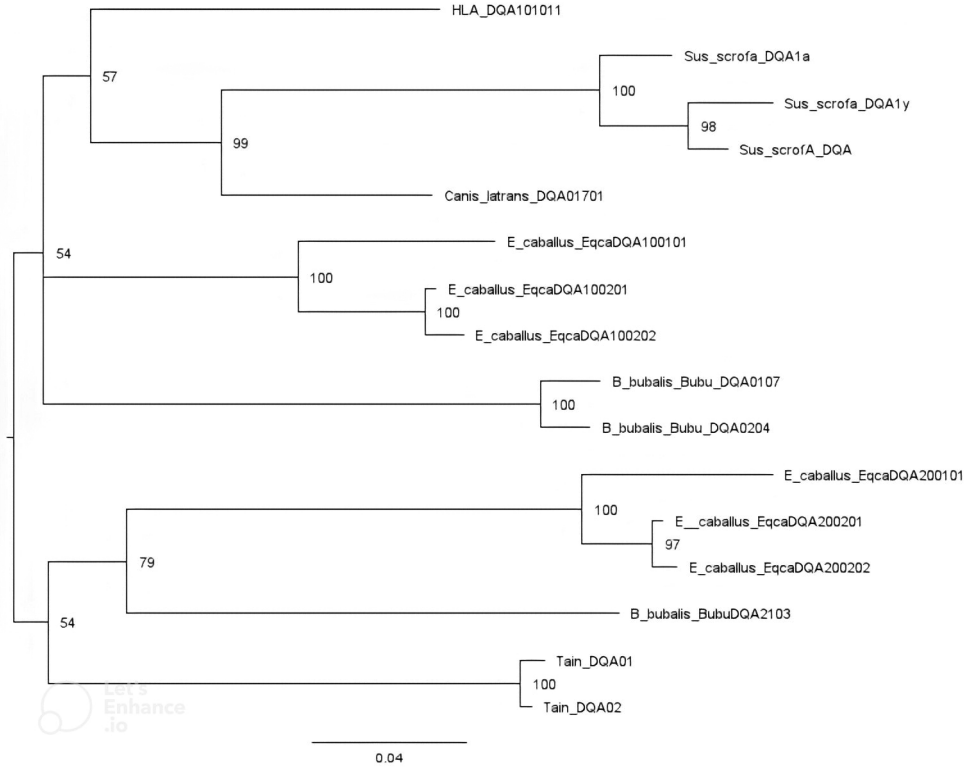

Fig. 4. Phylogenetic tree of MHC class II DQα domain (exon 2) sequences from Malayan tapir *Tapirus indicus*, equids, rhinoceros and other mammals including human bovine and canine. Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences in this tree are Tain_DQA01 and Tain_DQA02.

the DRA region shows the least within species and interspecies sequence variation when aligned with other closely and distantly related species. Higher amino acid variation could be observed within DRB, DQA and DQB genes sequences when aligned with other species. Although no more than two alleles for all class II genes were found in all individuals, up to three DRB loci are commonly found in mammalian species (Becker et al. 2009; Sin et al. 2012b). In contrast, studies in multiple species show the common lack of variation within the DRA region as compared to other class II genes. Many mammals have only one DRA, DQA, and DQB locus with a small number of alleles assigned to each loci (Takahashi et al. 2000).

All α and β class II genes in Malayan tapir show very high amino acid sequence similarity within their respective genes (> 89%). Although high similarity could indicate similar loci, we could not rule out the possibility of multiple loci in class II genes, as the alleles could not be conclusively assigned to their loci. Amino acid variation in compared sequences was least observed in the DRA gene, which could be because the species used for the sequence comparison in DRA were much more closely related *i.e.*, rhinoceros, than for the other regions and therefore, low variation is expected (Holmes and Ellis, 1999).

As shown in this study, the preliminary characterisation of the MHC gene from gDNA samples in Malayan tapir suggests that Malayan tapir possess orthologous gene loci consistent with the MHC gene in many vertebrates. The characterisation of this gene allele could reflect their function in producing MHC molecules as part of the immune response in this species. Therefore, Malayan tapir are also likely able to mount adaptive immune responses when there are pathogenic threats. However, the number of alleles found is relatively low compared to many MHC studies (Robinson et al. 2014). For example there are 8 alleles from two class I genes and 52 alleles from 8 class II genes in horses. This could be because the number

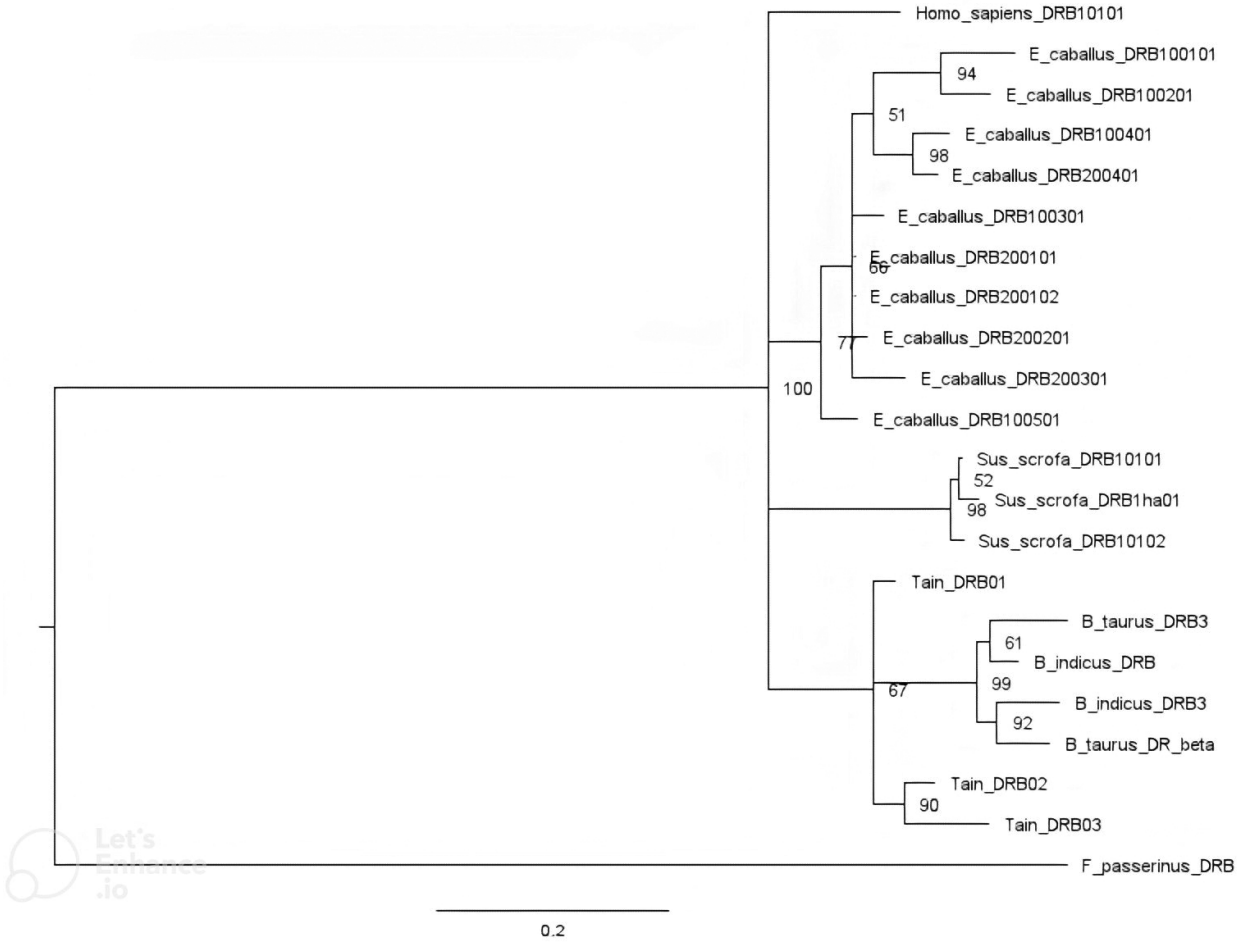

Fig. 5. Phylogenetic tree of MHC class II DRβ domain (exon 2) sequences from Malayan tapir *Tapirus indicus*, equids and other mammals including human and bovine. Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences in this tree are Tain_DRB01, Tain_DRB02, and Tain_DRB03.

of samples that we were able to get access to were only from captive animals and very limited due to the Endangered status of the studied species. Therefore, more studies need to be done to elucidate the Malayan tapir's MHC variation in wild populations.

Selection Analysis

Genetic variation within the MHC gene is generally attributed to a number of different processes which includes point mutation, recombination and gene conversion (Yeager and Hughes 1999; Miller and Lambert 2004). In the gene, point mutation acts on single nucleotides which give rise to new alleles. Fractions of gene conversion and recombination cause exchanges of nucleotide sequences between the allele and produce new haplotypes in the process. The variations within the genes that are generated by these processes are then maintained in a population by balancing selection that is primarily driven by parasite pressure in the population (Balasubramaniam et al. 2017). In class I MHC gene in Malayan tapir, the nucleotide variations within exon 2 and exon 3 are observed with higher non-synonymous substitutions over synonymous substitutions. The higher non-synonymous substitutions could be advantageous at the peptide binding region as non-synonymous substitutions essentially render amino acids and tend to be deleterious. The higher rates of non-synonymous over synonymous substitutions indicate evidence of balancing selection acting on the peptide-binding region (Hughes and Yeager 1998; Phillips et al. 2018). The evidence of selection is observed on the codon forming the antigen binding site (ABS). However, further analysis using PAML, could not detect positive selection sites (PSS) that acted by selection within the

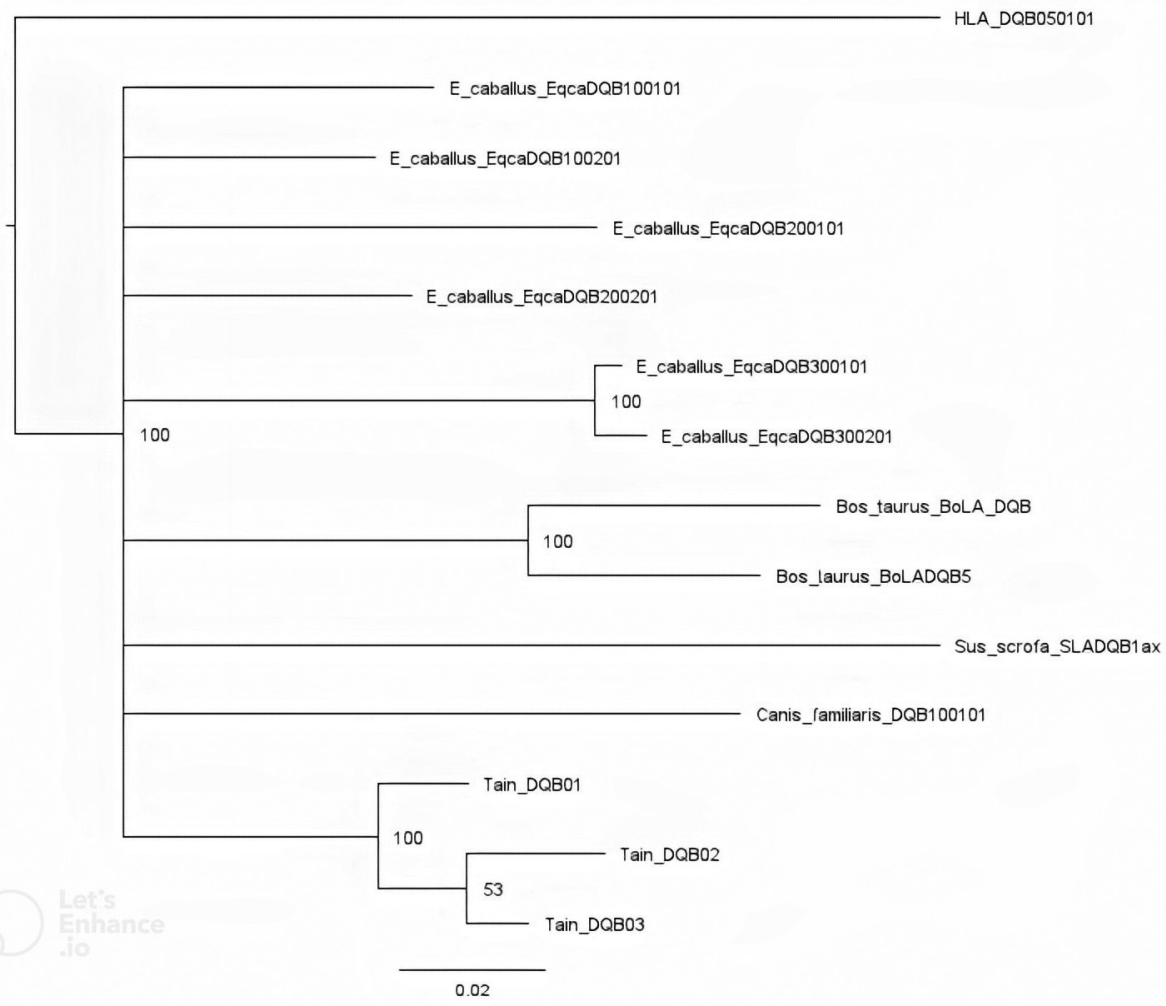

Fig. 6. Phylogenetic tree of MHC class II DQβ domain (exon 2) sequences from Malayan tapir *Tapirus indicus*, equids, rhinoceros and other mammals including human and canine. Bayesian posterior probabilities above 50% are shown above the branches. *Tapirus indicus* sequences in this tree are Tain_DQB01, Tain_DQB02, and Tain_DQB03.

class I exons.

In class II genes, evidence of selection could be observed in both the ABS and non-ABS regions in DRB, with a with higher value in the ABS region. Meanwhile, DQA, DQB, and DRA genes all are observed to not undergo selection at the peptide-binding region. The DRB in many species has also been observed to have an elevated rate of non-synonymous over synonymous substitutions in the ABS as compared to the other class II genes (Miller et al. 2011; Sin et al. 2012a b; Balasubramaniam et al. 2017). This could be because of the functional importance of this gene. Similarly, the other class II genes have also been observed to rarely display evidence of selection. Further analysis using PAML showed a significant result of comparing null and nested models of M1a vs M2a and M7 vs M8. Analysis using PAML concluded that 24 amino acids in DRB were detected as PSS, whereby 10 of the PSS in DRB regions were part of the amino acids forming the ABS, as determined in Reche and Reinherz (2003). This PSS in DRB supports the notion that balancing selection acts on the ABS and significantly contributes to the high polymorphism observed in exon 2 of the class II genes (Hughes and Yeager 1998; Osborne et al. 2015). This therefore implies the selective pressure shaping the variation in the class II genes in different mammals is driven similarly in Malayan tapir. For example, selective pressure is reflected in a relatively close family, Bovidae. Large numbers of cattle breeds across the world reported high numbers of polymorphism and allelic spectrum at the MHC-DRB3 locus, particularly at exon 2. The gene loci is also associated with a number of diseases such as resistance to dermatophilosis and higher susceptibility to subclinical mastitis depending on breed and alleles (Behl et al. 2012). Therefore, further studies on selection on the DRB gene and its potential association with disease in Malayan tapir could elucidate the mechanism that drives selection on the gene.

Phylogenetic Analysis

Malayan tapir have at least one class I gene and multiple class II genes. The gene sequences from *Tapirus indicus* seem to be grouped closer to rhinoceros species rather than horse species, although the distance between the supported clades are short. In class I exon 2 sequences, separate clustering of the sequences may indicate the presence of two loci. Exon 3 of class I regions cluster together except for Tain_C1_08 which may show closer relationships with other species. All class II genes show species specific monophyletic group formation except for DRB genes in their phylogenetic trees, in which case they form clusters with bovine species and interestingly show separate clustering to closely related horse species. The interspersed relationship indicated by the phylogenetic relationship within the class I and class II genes of Malayan tapir may indicate an occurrence of trans-species polymorphism of allelic lineage (Klein 1986). Trans-species alleles generally occurred in ancestral species and were retained over a long evolutionary timescale, even after the species diverged. Therefore, the MHC allele predating speciation event could be retained and passed on after speciation (Bernatchez and Landry 2003). The diverged species may also subsequently be subject to similar pathogen pressure, which therefore resulted in the retention of similar alleles. The interspersed relationships observed in the Malayan tapir sequences could also be results of concerted evolution, in which independent gene conversion transfers short stretches of DNA that continue to duplicate further and produce phylogenetic signals similar to that caused by trans-species polymorphism (Hess and Edwards 2002). Evidence of positive selection in Malayan tapir DRB gene coupled with the possibility of trans-species polymorphism could further support the idea that balancing selection may have occurred in the Malayan tapir population. However, more allele genotyping and samples are needed to better understand the process.

Limitations

Because of limited sample availability, the study could only focus on MHC alleles from gDNA. Therefore, studies on the functionality of those alleles need to be done based on RNA samples and also their level of expression. cDNA samples would be able to strongly indicate the functionality of the sequences obtained and to further analyse for evidence of selection. It is important to note that primers designed in this study could only capture a limited length. This results in partial length genes, and higher chances of missing out more sequences. Therefore, new primers should be designed and tested to capture longer lengths so that a more conclusive analysis can be made.

Importance of MHC gene study

In this study, we reported that the Malayan tapir has a low MHC diversity similar to other threatened species such as the giant panda (Zhu et al. 2007) and European mink (Becker et al. 2009). Thus, we should manage Malayan tapir populations both in captivity and in the wild to have a high variable MHC through artificial mate selection and translocation or reintroduction of dissimilar MHC to ensure the sustainability of population in the wild. Unrelated

individuals with maximal heterozygosity at the MHC or individuals with rare alleles are reported to be advantageous in many wild animals for their breeding success and survival from emerging new pathogens (Brandies et al. 2018; Han et al. 2019; Schwensow et al. 2019; Erofeeva et al. 2022). Quantifying MHC diversity within individuals and between populations are essential steps to identify potential improvements for captive management and conservation plans.

CONCLUSIONS

There are growing studies describing the structure and evolution of genes in non-model vertebrates. Our study is the first characterisation of MHC class I and II peptide-binding region in Malayan tapir. More importantly, this was done in Endangered large mammals that face severe population threats in Malaysia. The low population numbers increase concerns for inbreeding events and therefore might result in compromised genetic variation, particularly at the regions responsible for immune response *i.e.*, MHC gene. Increase in inbreeding reduces the allelic variation inherited and may cause the MHC molecules in inbred individuals to respond to only a limited number of pathogenic threats. Hence, mating of Malayan tapir in captivity could be matched with individuals of different populations to reduce the chance of inbreeding. Perhaps captivity centres could genotype the Malayan tapir at their MHC gene to avoid the mating of individuals with similar MHC alleles and variation. This would increase their chance of mounting appropriate recognition to wider pathogenic threats and increase their chances of survival. More individuals are also needed to be genotyped so that there is a clearer understanding on the allelic spectrum and polymorphism in the population.

Acknowledgments: We thank the management and veterinarians from Zoo Negara and Sg Dusun Wildlife Conservation Centre for sample collections and assistance in this study. This study took place under permit from the Ministry of Natural Resource and Environment [NRE 600-2/2/21 Jld.5(17)] and was supported by IPS Putra Grant (GP-IPS/2016/9505000), Universiti Putra Malaysia.

Authors' contributions: Ismail collected, ran experiments, analysed data and wrote up the draft. Yong designed parts of the methods and technical parts of the experiments. Sin reviewed the draft and finalised the results. Annavi supervised, reviewed and finalised the drafts.

Competing interests: The authors declare that they have no conflict of interest.

Availability of data and materials: The sequence data that support the findings of this study were assigned accession numbers and will be openly available in GenBank NCBI Nucleotide database at http://www.ncbi.nlm.nih.gov/nucleotide, with the following accession numbers MK432928-MK432945 and MK482362.

Consent for publication: The authors agree to the terms and conditions for publication.

Ethics approval consent to participate: All sampling and handlings procedures were approved by Institutional Animal Care and Use Committee, Universiti Putra Malaysia (Ethical approval ref.: UPM/IACUC/AUP-R033/2016).

REFERENCES

Akaike H. 1974. A new look at the statistical model identification. *In*: IEEE Transactions on Automatic Control **19:**716–723. doi:10.1109/TAC.1974.1100705.

Alberts B, Bray D, Hopkin K, Johnson AD, Lewis J, Raff M, Walter P. 2013. Essential Cell Biology (4th ed.). WW Norton & Company. doi:10.1201/9781315815015.

Albright-Fraser DG, Reid R, Gerber V, Bailey E. 1996. Polymorphism of DRA among equids. Immunogenetics **43:**315–317.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. doi:10.1016/s0022-2836(05)80360-2.

Balasubramaniam S, Mulder RA, Sunnucks P, Pavlova A, Melville J. 2017. MHC class II β exon 2 variation in pardalotes (Pardalotidae) is shaped by selection, recombination and gene conversion. Immunogenetics **69:**101–111. doi:10.1007/s00251-016-0953-7.

Barongi R. 1993. Husbandry and conservation of tapirs *Tapirus* spp. International Zoo Yearbook **32:**7–15. doi:10.1111/j.1748-1090.1993.tb03508.x.

Becker L, Nieberg C, Jahreis K, Peters E. 2009. MHC class II variation in the endangered European mink *Mustela lutreola* (L. 1761)—Consequences for species conservation. Immunogenetics **61:**281–288. doi:10.1007/s00251-009-0362-2.

Behl JD, Verma NK, Tyagi N, Mishra P, Behl R, Joshi BK. 2012. The major histocompatibility complex in bovines: a review. ISRN Vet Sci **2012:**872710. doi:10.5402/2012/872710.

Benton CH, Delahay RJ, Smith FA, Robertson A, McDonald RA, Young AJ, Hodgson D. 2018. Inbreeding intensifies sex- and age-dependent disease in a wild mammal. J Anim Ecol **87:**1500–1511. doi:10.1111/1365-2656.12878.

Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? J Evolution Biol **16:**363–377. doi:10.1046/j.1420-9101.2003.00531.x.

Blum JS, Wearsch PA, Cresswell P. 2013. Pathways of antigen processing. Annu Rev Immunol **31:**443–473. doi:10.1146/annurev-immunol-032712-095910.

Brandies PA, Grueber CE, Ivy JA, Hogg CJ, Belov K. 2018. Disentangling the mechanisms of mate choice in a captive koala population. PeerJ **6:**e5438. doi:10.7717/peerj.5438.

Cheng Y, Polkinghorne A, Gillett A, Jones EA, O'Meally D, Timms P, Belov K. 2018. Characterisation of MHC class I genes in the koala. Immunogenetics **70:**125–133. doi:10.1007/s00251-017-1018-2.

Ellis S, Martin A, Holmes E, Morrison W. 1995. At least four MHC class I genes are transcribed in the horse: Phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. Int J Immunogenet **22:**249–260. doi:10.1111/j.1744-313x.1995.tb00239.x.

Erofeeva MN, Alekseeva GS, Kim MD, Sorokin PA, Naidenko SV. 2022. Inbreeding coefficient and distance in MHC genes of parents as predictors of reproductive success in domestic cat. Animals **12:**1–16. doi:10.3390/ani12020165.

Fraser DG, Bailey E. 1998. Polymorphism and multiple loci for the horse DQA gene. Immunogenetics **47:**487–490. doi:10.1007/s002510050387.

Graur D, Li WH. 2000. Fundamentals of molecular evolution. Sinauer Associations Inc, Sunderland, MA.

Hall TA. 1999. BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *In*: Nucleic acids symposium series **41:**95–98.

Han QH, Sun RN, Yang HQ, Wang ZW, Wan QH, Fang SG. 2019. MHC class I diversity predicts non-random mating in Chinese alligators (*Alligator sinensis*). Heredity **122:**809–818. doi:10.1038/s41437-018-0177-8.

Hedrick P, Miller P. 1994. Rare alleles, MHC and captive breeding. Conserv Genet **68:**187–204. doi:10.1007/978-3-0348-8510-2_16.

Hess CM, Edwards SV. 2002. The Evolution of the Major Histocompatibility Complex in Birds: Scaling up and taking a genomic approach to the major histo compatibility complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the mhc. Bioscience **52:**423–431. doi:10.1641/0006-3568(2002)052[0423:teotmh]2.0.co;2.

Holmes E, Ellis S. 1999. Evolutionary history of MHC class I genes in the mammalian order Perissodactyla. J Mol Evol **49:**316–324. doi:10.1007/pl00006554.

Hughes AL, Hughes MK. 1995. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics **42:**233–243. doi:10.1007/bf00176440.

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.

Janeway CA, Travers P, Walport M, Schlomchik M. 2001. Immunobiology: The immune system in health and disease. Vol. 2, Garland Pub. New York.

Jukes TH, Cantor CR. 1969. Evolution of Protein Molecules. *In*: Munro, H.N., Ed., Mammalian Protein Metabolism, Academic Press, New York, pp. 21–132. doi:10.1016/B978-1-4832-3211-9.50009-7.

Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. Trends Ecol Evol **17:**230–241. doi:10.1016/S0169-5347(02)02489-8.

Klein J. 1986. Natural history of the major histocompatibility complex. Jan Klein, John Wiley and Sons: New York, 775 pages.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol **33:**1870–1874. doi:10.1093/molbev/msw054.

Kurtz BM, Singletary LB, Kelly SD, Frampton AR. 2010. *Equus caballus* major histocompatibility complex class I is an entry receptor for equine herpesvirus type 1. J Virol **84:**9027–9034. doi:10.1128/jvi.00287-10.

Li WH. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol **36:**96–99. doi:10.1007/BF02407308.

Lynam A, Traeholt C, Martyr D. 2008. *Tapirus indicus*. The IUCN Red List of threatened species 2011. Downloaded on 5 Oct. 2015

Miller HC, Bowker-Wright G, Kharkrang M, Ramstad K. 2011. Characterisation of class II B MHC genes from a ratite bird, the little spotted kiwi (*Apteryx owenii*). Immunogenetics **63:**223–233. doi:10.1007/s00251-010-0503-7.

Miller HC, Lambert DM. 2004. Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Mol Ecol **13:**3709–3721. doi:10.1111/j.1365-294x.2004.02368.x.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:**418–426. doi:10.1093/oxfordjournals.molbev.a040410.

O'Brien SJ, Evermann JF. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. Trends Ecol Evol **3:**254–259. doi:10.1016/0169-5347(88)90058-4.

Osborne AJ, Pearson J, Negro SS, Chilvers BL, Kennedy MA, Gemmell NJ. 2015. Heterozygote advantage at MHC *DRB* may influence response to infectious disease epizootics. Mol Ecol **24:**1419–1432. doi:10.1111/mec.13128.

Penn DJ. 2002. The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. Ethology **108:**1–21. doi:10.1046/j.1439-0310.2002.00768.x.

Phillips KP, Cable J, Mohammed RS, Herdegen-Radwan M, Raubic J, Przesmycka KJ, van Oosterhout C, Radwan J. 2018. Immunogenetic novelty confers a selective advantage in host-pathogen coevolution. Proc Natl Acad Sci USA. doi:10.1073/pnas.1708597115.

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 **331:**623–641. doi:10.1016/s0022-2836(03)00750-2.

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. 2014. The IPD and IMGT/HLA database: Allele variant databases. Nucleic Acids Res **43:**D423–D431. doi:10.1093/nar/gku1161.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572–1574. doi:10.1093/bioinformatics/btg180.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496–2497. doi:10.1093/bioinformatics/btg359.

Schwensow N, Castro-Prieto A, Wachter B, Sommer S. 2019. Immunological MHC supertypes and allelic expression: How low is the functional MHC diversity in free-ranging Namibian cheetahs? Conserv Genet **20:**65–80. doi:10.1007/s10592-019-01143-x.

Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T. 2012a. Evolution of MHC class I genes in the European badger (*Meles meles*). Ecol Evol **2:**1644–1662. doi:10.1002/ece3.285.

Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T. 2012b. 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.

Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool **2:**16. doi:10.1186/1742-9994-2-16.

Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. P Natil Acad Sci-Biol **101:**15261–15264. doi:10.1073/pnas.0403809101.

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci **277:**979–988. doi:10.1098/rspb.2009.2084.

Takahashi K, Rooney A, Nei M. 2000. Origins and divergence times of mammalian class II MHC gene clusters. J Hered **91:**198–204. doi:10.1093/jhered/91.3.198.

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.

Thompson JD, Gibson TJ, Higgins DG. 2003. Multiple sequence alignment using ClustalW and ClustalX. Current Protocols in Bioinformatics **1:**2–3. doi:10.1002/0471250953.bi0203s00.

Traeholt C, Novarino W, Bin Saaban S, Shwe NM, Lynam A, Zainuddin Z, Simpson B, Bin Mohd S. 2016. *Tapirus indicus*. The IUCN Red List of Threatened Species 2016:e. T21472A45173636. Downloaded on 18 July 2016.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res **35:**71–74. doi:10.1093/nar/gkm306.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:**1586–1591. doi:10.1093/molbev/msm088.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol **22:**1107–1118. doi:10.1093/molbev/msi097.

Yeager M, Hughes AL. 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev **167:**45–58. doi:10.1111/j.1600-065x.1999.tb01381.x.

Zhu L, Ruan XD, Ge YF, Wan QH, Fang SG. 2007. Low major histocompatibility complex class II DQA diversity in the Giant Panda (*Ailuropoda melanoleuca*). BMC Genet **8:**1–7. doi:10.1186/1471-2156-8-29.

Supplementary Materials

Fig. S1. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class I exon 2 clones, rhinoceros (*Diceros bicornis*), equids (*Equus caballus*), human (*Homo sapiens*) and cattle (*Bos taurus*). The GenBank accession numbers for α1 (exon2) sequences from other mammals are AF055346 (Diceros_bicornis_DibiUA01), DQ083407 (Equus_caballus_Eqca100101) and DQ145597 (Equus_caballus_Eqca100201), GU812295 (Homo sapiens_HLA_A01010101) and L02834 (Bos_taurus_classIBoLA). *Numbers above the sequence* indicate the codon position in the α1 domain. *Single letters* and *dots* represent amino acids that are distinct from or identical to Tapirus indicus sequence for this domain respectively. *Dashes* indicate missing sequences. Putative ABSs were defined according to Reche and Reinherz (2003) and are marked with an asterisk mark above the sequence. (download)

Fig. S2. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class I exon 3 clones, rhinoceros (*Rhinoceros unicornis Diceros bicornis*, and *Ceratotherum simum*), equids (*Equus caballus*), and cattle (*Bos taurus*). The GenBank accession numbers for α2 (exon3) sequences from other mammals are AJ133670 (R_unicornis_classI), AJ055348 (D_bicornis_DibiUB02), XM_014795072 (Ceratotherium_simum_classi), DQ083407 (E_caballus_Eqca100101) and DQ083408 (E_caballus_Eqca200101), and L02834 (Bos_taurus_classI_BoLA). *Numbers above the sequence* indicate the codon position in the α2 domain. *Single letters* and *dots* represent amino acids that are distinct from or identical to *Tapirus* indicus sequence for this domain respectively. *Dashes* indicate missing sequences. Putative ABSs were defined according to Reche and Reinherz (2003) and are marked with an asterisk mark above the sequence. (download)

Fig. S3. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class II exon 2 DRA clones, available Malayan tapir GenBank's sequence (*Tapirus indicus*), Baird's tapir (*Tapirus bairdii*), rhinoceros (*Diceros bicornis*, *Rhinoceros unicornis* and *Ceratotherum simum*), and equids (*Equus caballus*). The GenBank accession numbers for DRA (exon2) sequences from other mammals are KM347953 (T_indicus_Tain-DRA-0104) and KM347956 (T_indicus-Tain-DRA-0106), AF113547 (T_bairdii_Taba-DRA-0101), AF113549 (D_bicornis_DRA-0101), AF113554 (R_unicornis_DRA-0501), AF113553 (Cera_simum_DRA-0401), and JQ254081 (E_caballus_Eqca-DRA-00102). *Numbers above the sequence* indicate the codon position in the DRα domain. *Single letters*

and *dots* represent amino acids that are distinct from or identical to *Tapirus indicus* sequence for this domain respectively. *Dashes* indicate missing sequences. *Putative ABSs* were defined according to Reche and Reinherz (2003) and are marked with asterisk mark above the sequence. (download)

Fig. S4. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class II exon 2 DQA clones, human (*Homo sapiens*), equids (*Equus caballus*), bovine (*Bubalus bubalis*), boar (*Sus scrofa*), and coyote (*Canis latrans*). The GenBank accession numbers for DQA (exon2) sequences from other mammals are L3402 (HLA_DQA101011), JQ254060 (E_caballus_EqcaDQA100101), JQ254067 (E_caballus_EqcaDQA200202), KT428703 (B_bubalis_BubuDQA2103), AY285931 (Sus_scrofa_DQA1y) and AY126647 (Canis_latrans_DQA01701). *Numbers above the sequence* indicate the codon position in the DQα domain. *Single letters* and *dots* represent amino acids that are distinct from or identical to *Tapirus indicus* sequence for this domain respectively. *Dashes* indicate missing sequences. Putative ABSs were defined according to Reche and Reinherz (2003) and are marked with asterisk mark above the sequence. (download)

Fig. S5. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class II exon 2 DRB clones, equids (*Equus caballus*) and human (*Homo sapiens*). The GenBank accession numbers for DRB (exon2) sequences from other mammals are JQ254085 (E_caballus_DRB100101), JQ254084 (E_caballus_DRB100201), JQ254086 (E_caballus_DRB100301), and AF029288 (Homo_sapiens_DRB10101). *Numbers above the sequence* indicate the codon position in the DRβ domain. *Single letters* and *dots* represent amino acids that are distinct from or identical to *Tapirus indicus* sequence for this domain respectively. *Dashes* indicate missing sequences. Putative ABSs were defined according to Reche and Reinherz (2003) and are marked with asterisk mark above the sequence. (download)

Fig. S6. Amino acid sequence identity for the Malayan tapir *Tapirus indicus* class II exon 2 DQB clones, human (*Homo* sapiens), equids (*Equus caballus*), bovine (*Bos taurus*), boar (*Sus scrofa*) and dog (*Canis familiaris*). The GenBank accession numbers for DQB (exon2) sequences from other mammals are L34101 (HLA_DQB050101), JQ254070 (E_caballus_EqcaDQB100101), JQ254069 (E_caballus_EqcaDQB100201), DQ093609 (Bos_taurus_BoLa_DQB), AY459300 (Sus_Scrofa_SLADQB1ax), and AF016905 (Canis_familiaris_DQB10010). *Numbers above the sequence* indicate the codon position in the DQβ domain. *Single letters* and *dots* represent amino acids that are distinct from or identical to *Tapirus* indicus sequence for this domain respectively. *Dashes* indicate missing sequences. Putative ABSs were defined according to Reche and Reinherz (2003) and are marked with asterisk mark above the sequence. (download)

Table S1. Details of individuals included in this study. (download)

Table S2. Akaike Information Criterion (AIC) values for Malayan tapir MHC genes alignments with other species for phylogenetic analysis model selection. Model with lowest AIC value was selected for phylogenetic construction. lnL = log-Likelihood value. (download)