Molecular characterization of MHC class I genes in Four species of Turdidae family to assess genetic diversity and selection.

Muhammad Usman Ghani[1], Shakeel Hussain[1], An Buyang[1], Xu Yanchun[1], and Li Bo[2]

[1]Northeast Forestry University
[2]Affiliation not available

October 8, 2020

Abstract

In vertebrate animals, the molecules encoded by major histocompatibility complex (MHC) genes play an essential role in the adaptive immunity. MHC class I deal with intracellular pathogens (Virus) in birds. MHC class I diversity depends on the consequence of local and global environment selective pressure and gene flow. Here, we evaluated the MHC class I gene in four species of Turdidae family from a broad geographical area of northeast China. We isolated 77 MHC class I sequence, including 47 putatively functional sequences and 30 pseudo sequences from 80 individuals. Method based on analysis of cloned amplicons (n=25) for each species, we found two and seven MHC I sequence per individual indicates more than one MHC I loci identified in all sampled species. Results revealed an overall elevated genetic diversity at MHC class I, evidence of different selection patterns among the domains of PBR and non-PBR. Alleles found to be divergent with overall polymorphic sites per species ranging between 58 to 70 (out of 291 sites). Moreover, trans-species alleles were evident due to convergent evolution or recent speciation for the genus. Phylogenetic relationships among MHC I show an intermingling of alleles clustering among Turdidae family rather than between other passerines. Pronounced MHC I gene diversity is essential for existence of species. Our study signifies a valuable tool for the characterization of evolutionary relevant difference across a population of birds with high conservational concerns.

Molecular characterization of MHC class I genes in Four species of *Turdidae* family to assess genetic diversity and selection.

Muhammad Usman Ghani[1], Shakeel Hussain[2], Anboyang[3], Yan Chun Xu[4] and Dr Li Bo[5]*

[1]College of Wildlife Resources and Protected Area, Northeast Forestry University, Harbin 150040, China; drblouch18@gmail.com

[2]College of Wildlife Resources and Protected Area, Northeast Forestry University, Harbin 150040, China; shakil-alhusaini@hotmail.com

[3]Department of stem cell biology and medicine, Graduate school of medical science, Kyushu University, Fukuoka 810-0000, Japan; anboyang@med.kyushu-u.ac.jp

[4]State Forestry and Grassland Administration Research Center of Engineering Technology for Wildlife; College of Wildlife Resources and Protected Area, Harbin 150040, China; xu_danial@163.com

[5]College of Wildlife Resources and Protected Area, Northeast Forestry University, Harbin 150040, China; LIBO_770206@126.com

*Correspondence: drblouch18@gmail.com;Tel:(+92-333-602-2399), LIBO_770206@126.com; Tel: (+86-180-451-60508).

1

Abstract: In vertebrate animals, the molecules encoded by major histocompatibility complex (MHC) genes play an essential role in the adaptive immunity. MHC class I deal with intracellular pathogens (Virus) in birds. MHC class I diversity depends on the consequence of local and global environment selective pressure and gene flow. Here, we evaluated the MHC class I gene in four species of *Turdidae* family from a broad geographical area of northeast China. We isolated 77 MHC class I sequence, including 47 putatively functional sequences and 30 pseudo sequences from 80 individuals. Method based on analysis of cloned amplicons ($n=25$) for each species, we found two and seven MHC I sequence per individual indicates more than one MHC I loci identified in all sampled species. Results revealed an overall elevated genetic diversity at MHC class I, evidence of different selection patterns among the domains of PBR and non-PBR. Alleles found to be divergent with overall polymorphic sites per species ranging between 58 to 70 (out of 291 sites). Moreover, trans-species alleles were evident due to convergent evolution or recent speciation for the genus. Phylogenetic relationships among MHC I show an intermingling of alleles clustering among *Turdidae* family rather than between other passerines. Pronounced MHC I gene diversity is essential for existence of species. Our study signifies a valuable tool for the characterization of evolutionary relevant difference across a population of birds with high conservational concerns.

Keywords: MHC class I; *Turdidae* ; Adaptive immunity; Gene flow; Genetic diversity; Polymorphic; Convergent evolution

1 INTRODUCTION

Major Histocompatibility Complex (MHC) are molecules encoded by certain genes that are most polymorphic to have been described in vertebrates' genomes [1]. Two types of MHC gene families, class I and class II, which are useful cell surface glycoproteins that regulate the immune response. MHC class II molecules are heterodimers consists on α chain and a β chain, both contribute to presenting peptides from the processing of extracellular pathogens such as bacteria to the CD4+ T-helper cells [2]. Heterodimers molecules of MHC class I are made up of an α chain and a non-MHC molecule, the β2 microglobulin. The α chain constitutes a cytoplasmic tail, a transmembrane domain, and three extracellular domains named α1, α2 and α3 [3] that are encoded by exons 2, 3 and 4. The MHC class I molecules are expressed in almost all somatic cells and trigger an adaptive immune response by presenting endogenously derived peptides of viral protein and an individual's own body cells to CD8+ cytotoxic T-cells [4]. Polymorphism is largely confined within the region encoding the ABS (Antigen Binding Site) of the MHC class I [5]. Maintenance of surprising diversity is supposed to be take place by two types of selection: heterozygote advantage and frequency-dependent selection. Heterozygotes could recognize a broader range of antigens from multiple pathogens and therefore have more fitness than either individual having homozygote [6]. Other is frequency-dependent selection, in which rare alleles deliver a selective advantage where pathogens have found a means to escape against common immune defensive alleles in the population. Thus alteration in pathogens community with time and locality result in MHC variation in the host population. Generally, an individual possessing huge numbers and diverse MHC alleles; more pathogens it can recognize [1].

Structural diversity and immune response has been explored in numerous research, including genomics [7, 8], ailment [9-11], mate choice [12-14]. Sequence similarity at PBR based assignment to the locus is frequently hampered by various evolutionary indicators due to current recombination, duplication and/or concerted evolution as well as positive selection mediated by a variety of pathogens [15]. Thus, numerous studies emphasized MHC genes as important markers to evaluate the adaptive potential and evolutionary status of threatened population [16].

The emerging scenario inspires researchers to collect statistics from a group of wild taxa to enlarge our understanding of the evolution of MHC gene [17]. Despite significant efforts, protocols for locus-specific MHC genotyping in avian are still difficult to achieve and remarkably rare [18]. MHC studies in population of wild birds remain neglected possibly due to complications in amplifying gene sequences from bird species not closely related to systematically studied chicken [19, 20].

A significant decline in habitats and fragmentation of available habitats are predisposing factors for dramatic

deterioration in population sizes [21].The avian genus *Turdus* is one of the broadly distributed passerine genera, with 65 documented extant species. The genus is listed wild territorial birds that are beneficial to china having economic and research value. Birds of this genus are strongly migratory thus experiencing a variety of environments. Up to the present, there are no studies on MHC class I genes in *Turdidae*species, which is the first step towards exploring the role of selection mediated by pathogens in the maintenance of MHC class I diversity. Precisely, this study aims to (1) characterize locus-specific variation in MHC I exon 3 genes across the *Turdidae* family and to examine the mode of evolution by which such variation comes about. To achieve this, we have measured the diversity and selection at MHC I genes to make available the variations that exist across *Turdidae* family. (2) We investigate the numbers of alleles possessed by each species and the general features of alleles in terms of functional genetic diversity. (3) Phylogenetic analyses to assess evolutionary relationships and processes driving avian MHC I diversity among four species of *Turdidae* family and other avian species. Results of this study provide us an insight into the evolutionary processes involved in avian MHC diversity, an important path for detailed studies on avian MHC evolution.

2 MATERIAL AND METHODS

2.1 Study Population

The study population was non-sympatrically distributed 80 individuals of four species of genus *Turdus* of the *Turdidae* family. Samples include two to three contour feathers, tissue from breast and liver of birds accidentally injured or died during migratory season of 2017-19 in autumn and deposit in State key laboratory of wildlife detection center in northeast forestry university, stored at 4 °C. The sample material, voucher identity, and geographical location are presented (Table.S1; Figure 1).
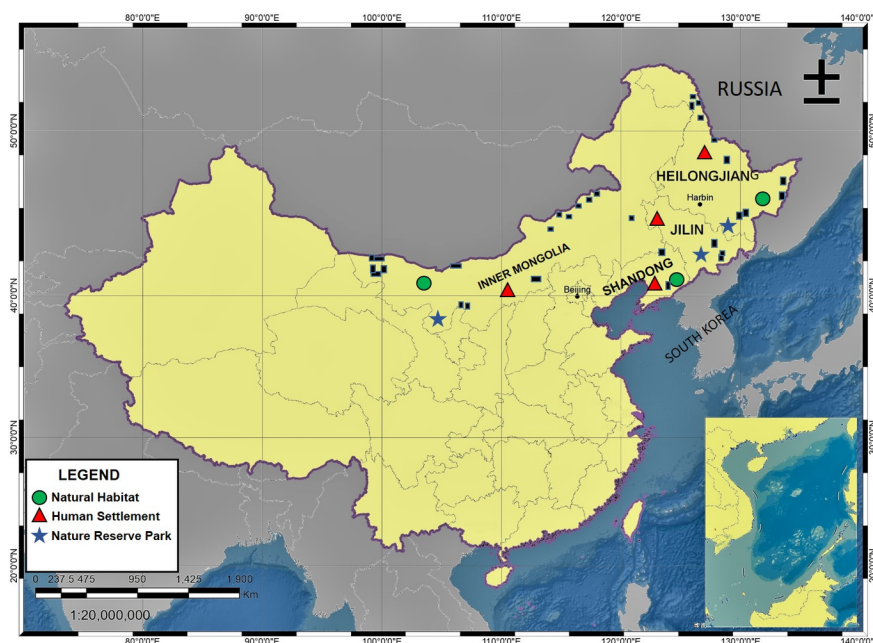
FIGURE 1: Geographical locations of samples included in our study. Square represents actual site of sample and size of square represents the approximate diameter of the sample's geographical range.

2.2 Extraction of genomic DNA

Region of calamus to the rachis of contour feathers was excised, tissues from skeletal muscles were minced, placed into a 1.5 ml Eppendorf tube containing TNE buffer (10mM Tris-HCl (pH 8.0), 150mM NaCl, 2mM EDTA, 1% SDS). Total genomic DNA was extracted with AxyPrep Multisource Genomic DNA Miniprep

3

Kit (AXYGEN, China) according to the manufacturer's instructions. The DNA concentration was measured with Nanopore Spectrophotometer at 260 nm absorbance. Samples above 100ng/µl concentration were used for further analysis.

2.3 PCR, cloning and sequencing

Polymerase chain reaction was conducted using motif specific primers designed for the amplification of MHC class I genes in great reed warbler. The forward primers HN36 5'-TCCCCACAGGTCTCCACACAGT-3' and HN46 reverse 5'-ATCCCAAATTCCCACCCACCTT-3' correspond to exon 3 region in the flanking introns, the region coding most of the peptide-binding site in MHC molecules (subunit α2) [22-24]. The primers were used due to their successful amplification in many passerine species. A list of primers trialed is given in Table S2. Amplification was performed in the reaction mixture containing 20ng DNA template, 0.2µM of each primer, 25µl 2× EasyTaq® PCR SuperMix (+dye) (Trans, China), and water (deionized) to reach 50µl as final volume. Thermal cycling for MHC class I amplification began with one cycle at 94°C for 5 min, followed by 30 cycles of denaturation consisting of sequential steps of 94°C for 30s, 52°C for 30s, and 72°C for 30s, ending with a single extension step at 72°C for 5 min. Purification was carried out with AxyPrep^{TM}DNA Gel Extraction Kit in accordance with the manufacturer's protocol. Purified PCR product was cloned using pEASY ®-T5 Zero Cloning Kit containing Trans1-T1 Phage resistant chemically competent cells (Transgen Biotech). PCRs were performed for positive clones using M13 forward and reverse primers. Several colonies (20-25) per individual were selected and used as a template for sequencing directionally on an automatic sequencer (ABI PRISM 3730; Invitrogen Biotechnology Co. Ltd.).

2.4 Definition of Allele

Since few artifacts introduced during the recombination of PCR products in cloning [25, 26]. Amplification, cloning and sequencing were performed twice. Sequences were verified and referred to as an Allele; either minimum of three sequences have the same nucleotide composition or repeated in both events. The sequences which showed any deletion, insertion, or premature stop codons within exons were identified as presumed pseudogene sequence, and others were considered as putative functional allele (PFA) [27]. All sequences appropriate to our criteria have been deposited into the GenBank (Accession No: MN849308-54).

2.5 Data analysis

2.5.1 Sequence Analysis

Chromatogram signals of all sequencing were examined with chromas 2.2.6. Sequences without ambiguous signals were selected. Vector sequence from the MHC class I gene was removed using seqMan in the DNA-Star7.1 package. Sequence editing and organization were done with BioEdit [28]. Sequences were aligned individually and then altogether four sampled species using CLUSTAL X [29]. The unique alleles were named according to the nomenclature for MHC in non-human species [30]. NCBI BLAST [31] was used for sequences confirmation representing close identity to passerine species previously published MHC class I exon 3 sequences. Sequences having at least one stop codon (shift in the reading frame due to indels or non-sense sequences) were classified as pseudogenes. Based upon sequences found to be translatable, a minimum number of functional loci MHC class I was estimated using a conservational approach that all Loci from samples species' individual were in heterozygote state.

The average pairwise nucleotide distances (Kimura 2-parameter model - K2P), and the Poisson-corrected amino acid distances were calculated using MEGA7.0. Standard errors were obtained through 1000 bootstrap replicates. Haplotypes identification (Na), the average number of nucleotide differences (K), polymorphic sites (S) and nucleotide diversity (π) were measured by DnaSP 5.10 [32].

2.5.2 Inference of recombination:

Recombination can influence the outcomes of selection, we first tested recombination. Analyses were implemented for the nucleotide alignment of exon 3 in the Recombination Detection Program version 4 (RDP4). Several method, including RDP [33], GENECONV [34], Chimaera [35], MaxChi [36], BootScan [33], SiScan

4

[37], and 3Seq [38], were used to detect recombination events. In addition, the online GARD tool, provided by the Datamonkey webserver (http://www.datamonkey.org/), was used for recombination signals assessment [39].

2.5.3 Tests for Selection

For selection, we conduct a priori classification of peptide binding region (PBR) and non- peptide region upon inferred passerine PBR sequences [40, 41] homology sites with chicken MHC [42, 43] and human HLA [44]. The identification of sites subjected to selection in MHC class I Exon 3 was performed using various methods. The first standard selection test (Tajima's D, Fu & Li's F* and Fu & Li's D*) were calculated using DnaSP 5.0 [32]. Second method was the calculation of parameter (ω) for functional alleles. It was carried out an overall estimation of d_N/d_S of MHC class I Exon 3 and the other was codons comprising only PBR and non-PBR, which was calculated with MEGA 7.0 according to the Nei-Gojobori method [45] with the Jukes and Cantor correction. Standard error estimates were derived from 1000 bootstrap replicates. Z test of historical positive selection [46] was calculated in MEGA 7.0. Third, the Maximum likelihood implemented in codeml in PAML 4.9 was used for identification of sites involved in the positive selection, which are indicated where the ratio ω (d_N/d_S) larger than 1 [47]. Two different models corresponding ω were tested: M7 (beta), M8 (β and ω). To find whether the alternative model (M8) provided better fitter than the M7, we performed Likelihood ratio tests to compare twice the difference of the log-likelihood ratios ($2\Delta lnL$) using a distribution χ^2. PSSs in the M8 model was identified by PP more than 95% using the Bayes empirical Bayes procedure. Positively selected sites were verified at each codon site separately using many complementary approaches implemented in datamonkey (http://www.datamonkey.org/) [48] in addition to afore mention methods. Specifically, we used MEME [49], FEL, SALC [50] and FUBER [51]

2.5.4 Phylogenetic Analysis:

To assess the phylogenetic relationship, we construct two phylogenies (One for sampled species and other representing MHC class I sequences of related passerines plus sampled species) using Bayesian inference. We find the GTR+T nucleotide substitution model [52] that fits our data using MrModeltest [35] through the Akaike Information Criterion (AICc) [53]. Bayesian Markov chain Monte Carlo (MCMC) was run for two million generations and sampling every 1,000 generations to ascertain when log Likelihood reached stationary phase. The phylogenetic tree was summarized in MrBayes v3.1.2 [54] and the first 25% of the tree as burn-in was removed. Fig tree was used for visualization of the consensus tree. Exploration of relation between sampled species and related avian species, we conducted a maximum likelihood (ML) analysis with MEGA 7.0 [55]. The data were analyzed with the T92+G model. We conducted 1000 bootstrap replicates to estimate the support. Values greater than 75% were indicated in the ML phylogenetic trees. The species covered are mainly from *Passeridae, Acrocephalidae, Paridae, Motacillidae, Muscicapidae, Hirundinidae, Phylloscopidae, Fringillidae, Cardinalidae, Sturnidae* . Representative sequences from samples species and other selected for phylogenetic analysis were summarized in (See Table S3) with respective studied reference from GenBank.

To further identify allelic lineages among sampled species and related avian species, we conducted the Neighbor-Net algorithm in SplitsTree 4.14.8. Neighbor-Net networks were based on uncorrected P-distances and carried out 1000 bootstrap replicates to estimate nodal support. Nodal support values (> 75%) were displayed.

3 RESULTS

3.1 Characterization of Alleles

We successfully and selectively amplified MHC class I exon 3 genes across 80 individuals from four species of *Turdidae* family using HN36 and HN46 primers. An average of 22.7 clones per individual was sequenced. Sequences varied between 459 and 579 base pairs. The multiple sequence alignments of all sampled species were 411 base pair long. The final aligned MHC class I dataset included 285-291bp (Primers not include). Analysis of gDNA alignment revealed a total of 77 distinct Haplotypes/alleles including 47 PFA. Frequency, identity, distribution and the proposed name of haplotypes recovered from the individual of *Turdidae* family are given

5

in (Supplementary Material Table S4). Each sequence was confirmed to exhibit similarity (81%-93%) with earlier reported passerine MHC class 1 sequences based upon BLAST search. The numbers of PFA sequences found in a single individual ranged from one to five, indicating that one to three loci exist in three of the four species of *Turdidae* family. However, the number of putative functional alleles found in a single individual ranged from two to seven in *Turdus atrogularis*exhibiting two to four loci. Number of the individual tested, number of PFA and pseudogene retrieved, the minimum number of functional loci estimated is given in Table.1. Three alleles (*Tuna-MHCI*PFA* 05=*Tuen-MHCI*PFA* 09,*Tuna-MHCI*PFA* 07=*Tuen-MHCI*PFA* 02 and*Tuen-MHCI*PFA* 05=*Tuna-MHCI*PFA* 015) were shared among*Turdus naumanni* and *Turdus eunomus.* Two alleles (*Turu-MHCI*PFA* 05=*Tuat-MHCI*PFA* 02 and*Turu-MHCI*PFA* 09=*Tuat-MHCI*PFA* 08) were also detected among individuals of *Turdus ruficollis* and *Turdus atrogularis.*Interestingly, genotypes comprising of one allele were by far the most repeated (26.67%, 8/30), followed by genotypes comprising two (16.67%, 5/30) and four alleles (13.3%, 4/30) in the population of *Turdus naumanni* . Almost pattern was consistent in population of *Turdus eunomus* and *Turdus rufficollis* . Genotypes constituting one allele (23.3%, 7/30) were the most repeated followed by three (16.67%, 5/30) in *Turdus eunomus.* Genotypes comprising one allele (33.33%, 5/15) were repeated in the population of *Turdus rufficollis.* Allelic repetition was absent in population of

Turdus

atrogularis.

Of the 77 sequences, 30 were non-translatable due to indels or the presence of stop codons resulted changes in the reading frame. Sequences were thus presumed to be pseudogenes. The number of identified pseudogenes within the four species ranged between three to five in most individuals of study population and six of the thirteen pseudogene sequences were found to be identical in three individuals from the population of sampled species. We cannot ignore the likelihood that some of the identified pseudogene sequences may be due to PCR or sequencing artifacts, as such events would more often result in nonfunctional sequences. The nucleotide deletion result in loss of 3 amino acids was obvious in *Tuna-MHCI*PS* 07-9 and *Tueu-MHCI*PS* 01-04 and*Tueu-MHCI*PS* 08. Both nucleotide deletion, frame shift mutation and premature stop codons were detected in *Turu-MHCI*PS* 01,03 and*MHCI*PS* 09 at amino acid 33 encoding Exon 3. Loss of 3 amino acids was at position 78 was detected in *Tuat-MHCI*PS* 05 and*Tuat-* MHCI*PS06.

3.2 Analysis of genetic diversity

Overall we find an elevated genetic diversity (π) within exon 3 alleles repertoire among individuals of *Turdus atrogularis* was (0.151) than *Turdus enumus* (0.113). The average number of nucleotides difference (*K*) varied between (43.95) in *Turdus atrogularis* to (32.32) in *Turdus eunomus.*

TABLE 1 : Amplification success and genetic diversity within each of the four species of *Turdidae* family investigated. MHC class I Exon 3 gene size (L), the overall number of polymorphic sites per allele repertoire (S), and the average number of nucleotide differences (*k*), Nucleotide diversity π at all sites, PBR and non-PBR.

Species	L	S	K	π
Turdus naummani	285	64	33.7	0.118 0.211 0.091
Turdus enumus	285	58	32.32	0.113 0.183 0.079
Turdus ruficollis	291	65	35.28	0.121 0.247 0.087
Turdus atrogularis	291	70	43.95	0.151 0.309 0.093

3.3 Analysis of Recombination

The recombination detection program not only analyzes brake points but also identify parent sequences. We ran the test of recombination by pooling all putative functional alleles recovered from four species of*Turdidae* family. We only find one potential recombination event in *Tuna-MHCI*PFA06* in*Turdus nau-*

6

manni at two recombinant breakpoints at position 148 and 253. *Tuna-MHCI*PFA02* as major and *Tuna-MHCI*PFA011* minor parent. Likewise, a single recombination was significant in *Tueu-MHCI*PFA07*. We detected no recombination among other alleles. However, these recombinations were only significant in two out of seven tests and not consistent with recombination breakpoint identified by GARD, hence the results represent that overall recombination is not likely to have any prominent effects on tests for positively selected sites (Table.2). The recombination breakpoints identified by these two programs are often inconsistent, probably because they use different computational methods.

TABLE 2: Recombinants detected in Turdidae family MHC class I alleles; parent sequences, breakpoints detected by the recombination detection program (RDP) and the genetic algorithm for recombination detection (GARD); and the RDP analyses.

		Recombination event 1	Recombination event 2
	Recombinant	*Tuna_MHCI*PFA06*	*Tuna_MHCI*PFA02*
	Maj P	*Tuna_MHCI*PFA02*	*Tuna_MHCI*PFA09*
	Min P	*Tuna_MHCI*PFA011*	Unkonwn (*Tuna_MHCI*PFA01*)
	BP 1 location	148 (148)	254 (253); $P< 0.001$
	BP 2 location	Absent	Absent
RDP methods	RDP	NS	<0.05
	GENECONV	<0.001	<0.001
	BootScan	<0.01	<0.05
	MaxChi	<0.001	<0.01
	Chimaera	<0.01	<0.01
	SiScan	<0.001	<0.001
	3Seq	<0.001	<0.01
	BP from GARD	Absent	Absent

Note: NS indicates not significant. Maj P and Min P represent as major and minor parents. BP denote Breakpoint. The numbers in parentheses are BP locations in the recombinant nucleotide sequences without gaps. The values after the semicolon are Max Chi values for that BPs. **indicate $P < 0.01$.

Analysis of Selection

Considering that the evolutionary history of each domain might have been different, we tested each domain separately for evidence of positive selection. Selection statistics by traditional methods did not disclose any statistical significant signal of selection that deviate from neutral expectations for *turdus enumus* (Tajima's D: -0.87309, P > 0.10; Fu & Li's D* test statistic: 0.36, P > 0.10; Fu & Li's F* test statistic: 0.03, P > 0.10) and *turdus atrogularis* (Tajima's D: -0.86107, P > 0.10 Fu & Li's D* test statistic: 0.19, P > 0.10; Fu & Li's F* test statistic: -0.077, P > 0.10). Still, overall d_N value was significantly higher statistically than dS in *Turdus atrogularis* (1.687) and ratio $d_{N/}ds$ was more pronounced at codons presumably coding PBR (1.994) than codons not involved in such activity (0.884) is presented in (Table 3).

TABLE 3: The average rates of nonsynonymous (d_N), synonymous (ds) substitutions and the result of Z-test, the average nucleotide distances (d_{nt}) and amino acid distances (d_{aa}) for PBR, non-PBR and all sites in MHC class I of *Turdidae* family.

Specie	Domain	$d_N\pm$**SE**	$ds\pm$ **SE**	**Z**	**P**	ω	$d_{nt}\pm$**SE**	$d_{aa}\pm$**SE**
Turdus naummani	All sites	0.142±0.026	0.104±0.013	0.610	0.543	1.365	0.051± 0.08	0.246± 0.035
	PBR	0.281±0.059	0.153±0.032	1.848	**0.034**	1.835	0.041± 0.045	0.372± 0.082

Specie	Domain	$d_N\pm$SE	$d_S\pm$ SE	Z	P	ω	$d_{nt}\pm$SE	$d_{aa}\pm$SE
	Non-PBR	0.051±0.062	0.067±0.012	1.356	0.476	0.761	0.034± 0.09	0.126± 0.029
Turdus enumus	All sites	0.134±0.021	0.107±0.019	0.729	0.457	1.251	0.043± 0.010	0.202± 0.031
	PBR	0.179±0.023	0.102±0.032	1.442	0.383	1.175	0.057± 0.031	0.366± 0.011
	Non- PBR	0.056±0.041	0.049±0.012	1.792	1.000	1.142	0.051± 0.011	0.134± 0.029
Turdus ruficollis	All sites	0.146±0.048	0.101±0.025	0.911	0.771	1.445	0.059± 0.017	0.191± 0.035
	PBR	0.264± 0.037	0.138±0.011	1.643	**0.002**	1.912	0.147± 0.045	0.453± 0.078
	Non- PBR	0.070±0.018	0.068±0.028	0.061	0.476	1.029	0.034± 0.049	0.126± 0.052
Turdus atrogularis	All sites	0.189±0.091	0.112±0.025	1.040	0.150	1.687	0.063± 0.053	0.211± 0.039
	PBR	0.321±0.013	0.161±0.069	1.012	0.435	1.994	0.207± 0.045	0.572± 0.162
	Non- PBR	0.069±0.062	0.078±0.012	1.813	0.476	0.884	0.078± 0.091	0.206± 0.041

* The errors were attained through 1000 bootstrap replicates are in parentheses.

+ Bold represents significant results.

Application of Likelihood models represents that the model M8 allows for positive selection provides a better than the neutral evolution models M7. Sites being positively selected were recognized, are given in (Table 4). In total, we find 12 codons under positive selection in sampled species, of which three sites (25%) match homologues codons found positively selected in other avian species and one (8.3%) matched human peptide binding region (Table 4).

TABLE 4: Estimation of d_N and d_S substitution rates for sites positively selected and their ratio for codons chosen a priori to constitute (PBR) and not involved (non-PBR).

Species	Comparison	Model	lnL value	Parameter estimates	PSSs	LRT	TS value
Turdus naummani	Tuna 1-30	M7 (Beta) M8 (Beta & Omega)	-703.53 -722.49	P = 0.19, q =0.138, P_0 = 0.351 P = 0.87, q = 0.117, P_1 = 0.09, ω = 3.11	Not allowed *39F*,41L *88T*	M7 vs. M8	5.47
Turdus enumus	Tuna 1-30	M7 (Beta) M8 (Beta & Omega)	-692.13 -714.11	P = 0.13, q = 0.97, P_0 = 0.347 P = 0.91, q = 0.158, P_1 = 0.11, ω = 3.76	Not allowed 41L, *52P,88T*	M7 vs. M8	4.56
Turdus ruficollis	Turu1-15	M7 (Beta) M8 (Beta & Omega)	-811.27 -834.88	P = 0.49, q = 0.187, P_0 = 0.411 P = 0.131, q = 0.232, P_1 = 0.13, ω = 3.94	Not allowed *29Y*, **78F** , *88G*	M7 vs. M8	6.21

| Species	Comparison	Model	lnL value	Parameter estimates	PSSs	LRT	TS value
Turdus atrogularis	Tuat1-5	M7 (beta) M8 (Beta & Omega)	-847.53 -849.78	P = 0.83, q = 0.211, P_0 = 0.585 P = 0.173, q = 0.279, P_1 = 0.17, ω = 4.11	Not allowed *39H*, 78F, **91Q**	M7 vs. M8	7.10

*The log-likelihood values and parameters estimated were computed using codeml implemented in PAML 4.9. PSSs were inferred in model M8 by BEB with posterior probabilities > 95%.

Usually consistent with the above finding, every substitute test (MEME, SALC, FEL, and FUBAR) for positive selection implemented in online adoptive evolutionary server Datamonkey (Weaver et al., 2018) identify numerous codons under positive selection (Figure 2).
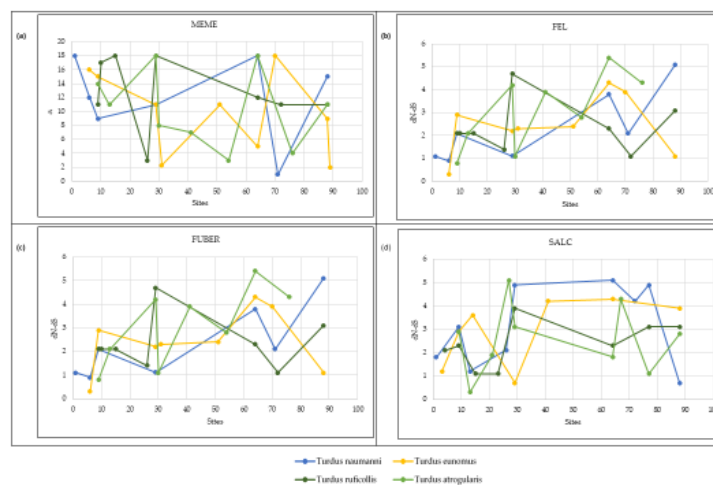
FIGURE 2: Positively selected sites using online adoptive evolutionary server Datamonkey with **(a)** MEME. **(b)** FEL.**(c)** FUBAR and **(d)** SALC. Substitution tests identify numerous codons showed the signature of positive selection.

Across all tests for positive selection, four codons (9,29,65 and 88) were frequently identified by all methods as having under positive selection. Of these, codons (42, 59) were corresponding to PBR in human and codons 9, 29, 64 and 88 also match homology to PBR, known as positively selected among passerine in general [56] (Figure 4). The ten most frequent MHC class I alleles retrieved from sampled species displayed 87%-91% sequence similarity to 18 sequences from five other passerine families*(Acrocephalidae, Passeridae, Muscicapidae, Paridae, Passerellidae)*. None of the 77 alleles studied had 100% sequence similarity to other published sequences to GenBank; thus, it establishes no allelic pair in the study population that was 100 % sequence likeness shared by another species.

9
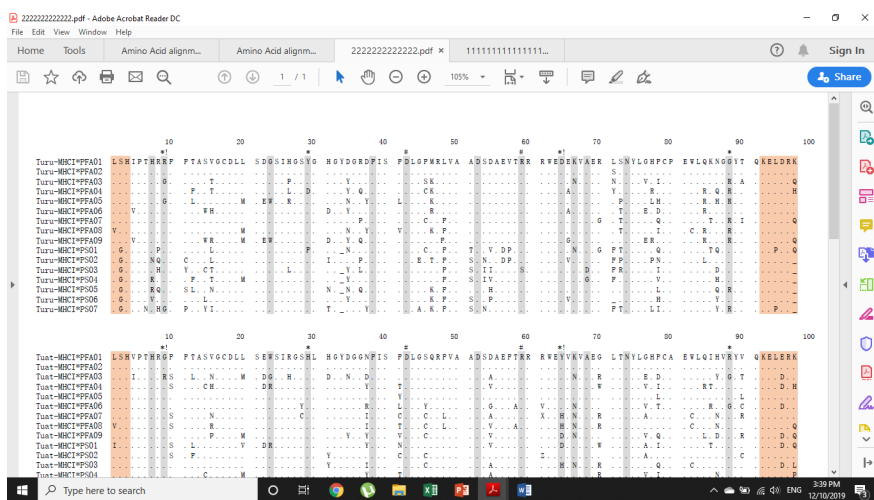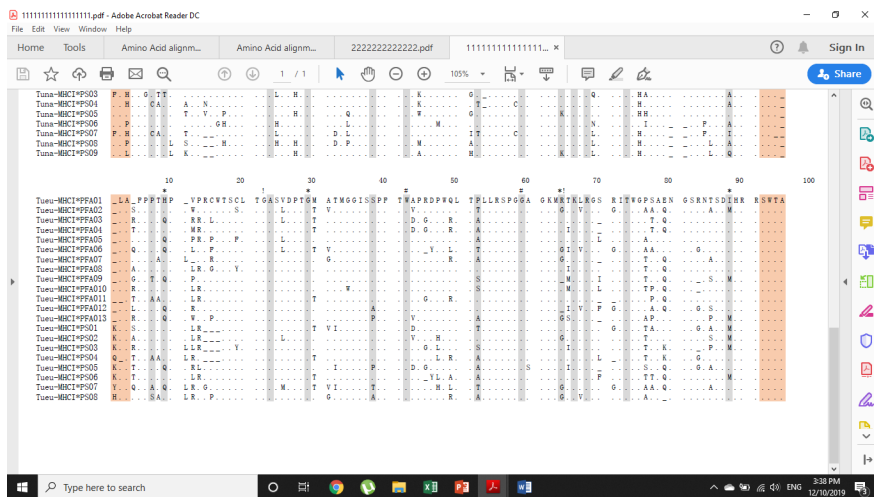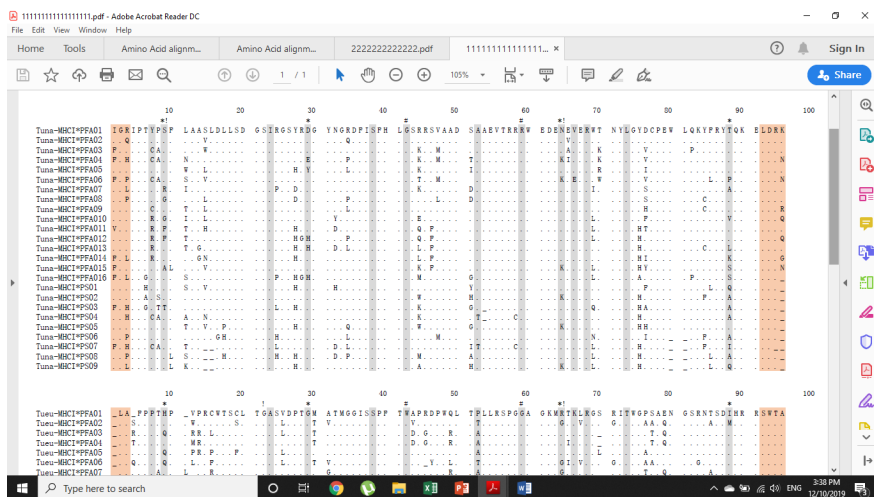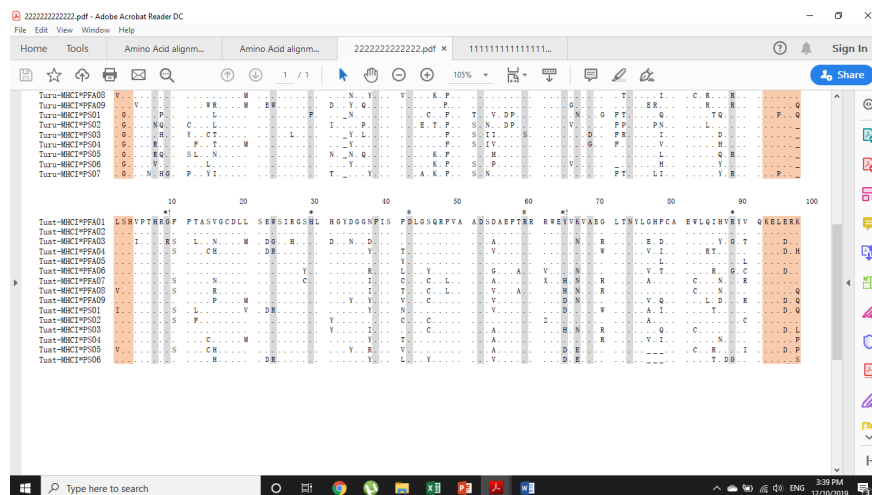
Figure 3: Alignment of deduced amino acid sequences of all allele retrieved from four species of *Turdidae* family. Dots (.) indicate identity with the reference sequence. Light gray color represents codons presumably coding for peptide binding regions upon alignment with human and other avian species. Light brown region indicate flanking introns. (-) indicate missing nucleotides. (∗) Represents codons positively selected in almost all of test perform for selection analysis. (!) indicates homologues region also positively selected in most of avian species. (#) Inferred homology with human.

3.5 Phylogenetic Analysis

In phylogenetic analysis, we observed that sampled species form a well-supported monophyletic clade with *Erithacus rubeculs* members of *Turdidae* family in maximum likelihood analysis (Figure S1). Bayesian analysis represents that most of the alleles shared among *Turdus atrogularis* and *Turdus reficollis* . This pattern was almost consistent among *Turdus naumanni* and *Turdus eunomus* presented in Figure 4. The Net network of putative functional and pseudogene MHC class I exon 3 sequences in *Turdidae* family with other passerines indicate that allelic distribution among them is almost congruent with limited divergence. For instance, *Tueu-MHCI*PFA02* and *Tuna-MHCI*PFA07* networks formed a monophyletic clade in the phylogenetic network of exon 3. Three alleles were shared among *Turdus naumanni* and *Turdus eunomus* two among *Turdus rufficollis* and *Turdus atrogularis.* The clustering of the sequences among species could be due to trans-species polymorphism or orthology [57].

FIGURE 4: Bayesian phylogenetic reconstruction of MHC class I exon 3 of four species of *Turdidae* family. All the nodes are well supported (PP>0.90%) unless indicated otherwise. AB268885 Gallus gallus was used as an outgroup.

DISCUSSION

In this study, we have for the first time characterize MHC Class I gene in four species of *Turdidae* family in the order Passeriformes from the wide geographical area of Northeast china. Analysis of MHC class I sequences revealed a total of 77 distinct Haplotypes/alleles including 47 putative functional alleles ever reported in passerine species, a group which is reported to have surprising MHC diversity [58, 59]. According to our findings based on MHC class I sequences, the functional loci in an individual ranged from one to three in three of the four species, which was consistent with findings from other passerine species studied till now [60]. In addition, we detected a large number of presumed pseudogene sequences in the sampled population as it retains important information about the evolution of MHC. This is not surprising, as it is consistent

with the expectation of evolution by birth-and-death [61]. We made a significant effort to characterize the variation in regions of MHC class I exon 3 in our study population, we find that the primers would make some unlikely bias in allelic variations among individuals. Hence, MHC class I alleles variations per individual should, largly be due to copy number of genes variation among individuals, which has been confirmed in other birds [62]. Few MHC class I alleles were shared between *Turdus naumanni* and *Turdus eunomus* as well as among individuals of *Turdus ruficollis* and *Turdus atrogularis* is indicating allelic sharing due to common ancestors or challenging common pathogens, as this event is frequent in numerous avian species such as owls, ardeid birds, penguins and passerines [63-65].

Generally, abundant variation in genetic material in a species is an indicator of the capacity to adapt to numerous environmental changes by that species. Rapidly evolving environmental pathogens would cause MHC genes to exhibit enlarged genetic diversity in species [66, 67]. Collectively, in our study, we find elevated genetic diversity among functional sequences and significant divergence, whereas pseudogene has low genetic variation and limited divergence. Similar results also have been described in other passerine species, including common yellow throat [68], great reed warbler and the great tit [69]. The allelic variation described in our study could be due to increased immunological defense against the internal pathogen since these are highly unlikely to adapt to novel, infrequent variant [15].

Recombination has been considered an important mechanism that influences allelic diversity and driving evolution of the MHC gene [70, 71] We only find one potential recombination event in *Tuna-MHCI*PFA06* at two recombinant breakpoints at position 148 and 253 identified with recombination detection program. Similarly, single recombination was significant in *Tueu-MHCI*PFA07*. Recombination pattern was also restricted two out of seven tests; hence our finding indicate recombination is unlikely to have any significant influence on tests for PSs. Though we could not find any substantial recombination among other alleles, qualitatively our result suggests a role for recombination during the evolution of MHC class I in our species studied. Our finding is consistent with, that micro-recombination is frequently observed in MHC genes [57]. Further study of recombinant function in the future will contribute to a detailed understanding of its role in the evolution of the MHC gene.

Positive selection is the maintainer of alleles having the advantageous mutation that maintain fitness of an individual. In our study, the classical test of selection Tajima's D, Fu & Li's D* and Fu & Li's F* showed no deviation from neutral selection or balance selection. Considering the level of variation, conventional methods used to find selection are not influential [72]. As sites positively selected are likely to accumulate more non-synonymous than synonymous substitutions, influencing amino acid variation to result in functional modifications in proteins [73]. Our study revealed differential expression of selection pattern in functional sequences on regions related with PBR and non-PBR of the MHC class I gene. Codons involved in peptide binding region revealed more non-synonymous substitution than synonymous (d_N/d_S=1.99) in *Turdus atrogularis* as compared to non- peptide binding region (d_N/d_S=0.884), pattern was consistent among all species tested, which might be enlightened that stronger selection pressure from intracellular pathogens than extracellular pathogens [74]. Evidence of positive selection at PBR of MHC has been reported in the house sparrow (PBR d_N/d_S = 1.55 vs non-PBR d_N/d_S = 0.51) [75] and golden pheasant (PBR d_N/d_S = 1.45 vs non-PBR d_N/d_S =0.91) [76]. Of the 12 codons in total among species tested exhibit positive selection with Likelihood methods using PAML, 9,29, 64 and 88 match homologues codons found positively selected in other passerine species.

It should be noted that the pooling of all alleles across loci will mostly reduce selection detection tests, so the outcomes might be conservative, but will be less prone to false positives [77, 78]. Therefore, attention should be given while inferencing about the detected diversity in MHC and the possible effects of selection on individual loci. Our results suggested that α2 domain of MHC class I exon 3 of all species are under positive selection pressure. Pronounced positive selection at antigen-binding sites permits a species or population to present a larger repertoire of peptides (antigens), thus increase the defensive ability against parasitic and pathogenic infections.

Finally, phylogenetic clustering of MHC class I data set of sampled species when pooled with other passerine

species produces a contrasting pattern. In general, the MHC class I sequence of *Turdidae* family clustered together with sequences from congeneric species. We found increased sequences similarities between same species rather than within species (trans specific likenesses), is usually described with trans species polymorphism (TSP), which occurs due to alleles passage from ancestral to the decedent via partial arrangement of lineages [79]. Although trans specific similarities can be described with convergent evolution due to the results of similar environmental selective pressure. Studies indicate that TSP is a primary mechanism responsible for clustering of alleles at avian MHC class I [80].
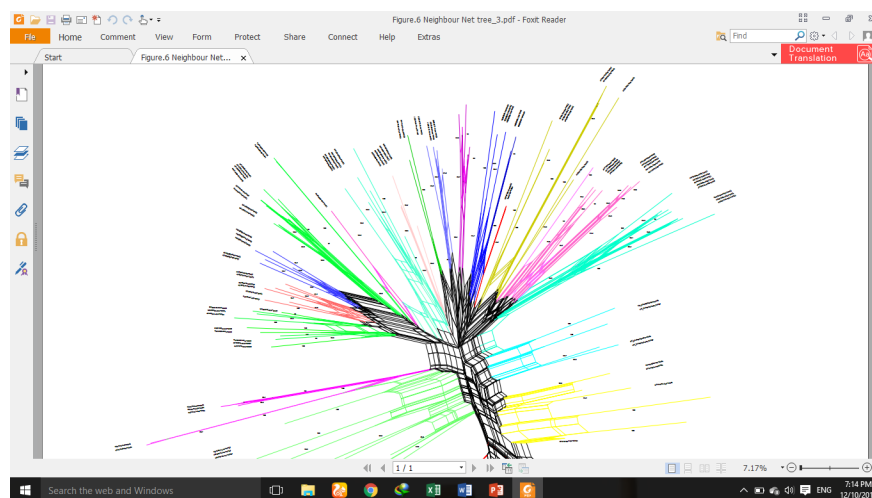
FIGURE 5: The phylogenetic networks of MHC class I of four species of turdidae family along with homologues sequences from passerines species. GenBank accession number are provided. Species name are mention lower right side with branch colors. Neighbor-Net networks based on uncorrected P-distances and carried out 1000 bootstrap replicates to estimate nodal support. Nodal support values ($> 75\%$) were displayed.

CONCLUSION

Our study shows that species of *Turdidae* family has retained significant MHC class I diversity, which supports high conservational value and contribute to evolution of MHC class I genes. Importantly, we specifically amplify the exon 3 locus, provide an opportunity to avoid chimera formation during molecular characterization of hypervariable genes of immunity. At the same time, our study is the first to validate contrasting patterns of allelic diversity and positive selection upon inferred PBR and non-PBR codons supported the hypothesis that different mechanisms can shape evolutionary paths of MHC class I.

Data availability statement:

Data will be archived and available through doi for assessing data in public and open access.

Conflicts of Interest: The authors declare no conflict of interest.

13

References

1. Murphy, K., P. Travers, and M. Walport, *Janeway's immunobiology. Garland science.* New York, 2008.

2. Jensen, K.K., et al., *Improved methods for predicting peptide binding affinity to MHC class II molecules.* Immunology, 2018. 154(3): p. 394-406.

3. Bjorkman, P.J. and P. Parham, *Structure, function, and diversity of class I major histocompatibility complex molecules.* Annual review of biochemistry, 1990. 59(1): p. 253-288.

4. Zhang, X.-H., et al., *Molecular characterization, balancing selection, and genomic organization of the tree shrew (Tupaia belangeri) MHC class I gene.* Gene, 2013. 522(2): p. 147-155.

5. Yeager, M. and A.L. Hughes, *Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution.* Immunological reviews, 1999. 167(1): p. 45-58.

6. Piertney, S. and M. Oliver, *The evolutionary ecology of the major histocompatibility complex.* Heredity, 2006. 96(1): p. 7.

7. Kulski, J.K., et al., *Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man.* Immunological reviews, 2002. 190(1): p. 95-122.

8. Kelley, J., L. Walter, and J. Trowsdale, *Comparative genomics of major histocompatibility complexes.* Immunogenetics, 2005. 56(10): p. 683-695.

9. Langefors, Å., et al., *Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2001. 268(1466): p. 479-485.

10. Hedrick, P.W., *Pathogen resistance and genetic variation at MHC loci.* Evolution, 2002. 56(10): p. 1902-1908.

11. Westerdahl, H., *Passerine MHC: genetic variation and disease resistance in the wild.* Journal of Ornithology, 2007. 148(2): p. 469-477.

12. Paterson, S. and J.M. Pemberton, *No evidence for major histocompatibility complex–dependent mating patterns in a free–living ruminant population.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 1997. 264(1389): p. 1813-1819.

13. Landry, C., et al., *'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar).* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2001. 268(1473): p. 1279-1285.

14. Knafler, G.J., et al., *MHC diversity and mate choice in the magellanic penguin, Spheniscus magellanicus.* Journal of Heredity, 2012. 103(6): p. 759-768.

15. Borghans, J.A., J.B. Beltman, and R.J. De Boer, *MHC polymorphism under host-pathogen coevolution.* Immunogenetics, 2004. 55(11): p. 732-739.

16. Jepson, A., et al., *Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens.* Infection and Immunity, 1997. 65(3): p. 872-876.

17. Edwards, S.V., et al., *A 39-kb sequence around a blackbird Mhc class II gene: Ghost of selection past and songbird genome architecture.* Molecular Biology and Evolution, 2000. 17(9): p. 1384-1395.

18. Canal, D., et al., *Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher Ficedula hypoleuca.* BMC research notes, 2010. 3(1): p. 236.

19. Kroemer, G., et al., *Molecular genetics of the chicken MHC: current status and evolutionary aspects.* Immunological reviews, 1990(113): p. 119-145.

20. Zoorob, R., et al., *Chicken major histocompatibility complex class II B genes: analysis of interallelic and inter-locus sequence variance.* European journal of immunology, 1993. 23(5): p. 1139-1145.

21. Peacock, M.M. and A.T. Smith, *The effect of habitat fragmentation on dispersal patterns, mating behavior, and genetic variation in a pika (Ochotona princeps) metapopulation.* Oecologia, 1997. 112(4): p. 524-533.

22. Bonneaud, C., et al., *Diversity of Mhc class I and IIB genes in house sparrows (Passer domesticus).* Immunogenetics, 2004. 55(12): p. 855-865.

23. Westerdahl, H., *No evidence of an MHC-based female mating preference in great reed warblers.* Molec-

ular Ecology, 2004. 13(8): p. 2465-2470.

24. Promerova, M., T. Albrecht, and J. Bryja, *Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (Carpodacus erythrinus).* Immunogenetics, 2009. 61(6): p. 451-461.

25. Babik, W., et al., *New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system.* Molecular ecology resources, 2009. 9(3): p. 713-719.

26. Kanagawa, T., *Bias and artifacts in multitemplate polymerase chain reactions (PCR).* Journal of bioscience and bioengineering, 2003. 96(4): p. 317-323.

27. Abduriyim, S., et al., *Evolution of MHC class I genes in Eurasian badgers, genus Meles (Carnivora, Mustelidae).* Heredity, 2019. 122(2): p. 205.

28. Hall, T.A. *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT* . in *Nucleic acids symposium series* . 1999. [London]: Information Retrieval Ltd., c1979-c2000.

29. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* bioinformatics, 2007. 23(21): p. 2947-2948.

30. Klein, J., et al., *Nomenclature for the major histocompatibility complexes of different species: a proposal* , in *The HLA system in clinical transplantation* . 1993, Springer. p. 407-411.

31. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. 215(3): p. 403-410.

32. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.* Bioinformatics, 2009. 25(11): p. 1451-1452.

33. Martin, D., et al., *A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.* AIDS Research & Human Retroviruses, 2005. 21(1): p. 98-102.

34. Padidam, M., S. Sawyer, and C.M. Fauquet, *Possible emergence of new geminiviruses by frequent recombination.* Virology, 1999. 265(2): p. 218-225.

35. Posada, D., *jModelTest: phylogenetic model averaging.* Molecular biology and evolution, 2008. 25(7): p. 1253-1256.

36. Smith, J.M., *Analyzing the mosaic structure of genes.* Journal of molecular evolution, 1992. 34(2): p. 126-129.

37. Gibbs, M.J., J.S. Armstrong, and A.J. Gibbs, *Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences.* Bioinformatics, 2000. 16(7): p. 573-582.

38. Boni, M.F., D. Posada, and M.W. Feldman, *An exact nonparametric method for inferring mosaic structure in sequence triplets.* Genetics, 2007. 176(2): p. 1035-1047.

39. Kosakovsky Pond, S.L., et al., *GARD: a genetic algorithm for recombination detection.* Bioinformatics, 2006. 22(24): p. 3096-3098.

40. Balakrishnan, C.N., et al., *Gene duplication and fragmentation in the zebra finch major histocompatibility complex.* BMC biology, 2010. 8(1): p. 29.

41. Alcaide, M., et al., *Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot Fulica atra (Aves: Rallidae).* Ecology and evolution, 2014. 4(6): p. 688-698.

42. Kaufman, J., J. Salomonsen, and M. Flajnik. *Evolutionary conservation of MHC class I and class II molecules—different yet the same* . in *Seminars in immunology* . 1994. Elsevier.

43. Hee, C.S., et al., *Structure of a classical MHC class I molecule that binds "non-classical" ligands.* PLoS biology, 2010. 8(12): p. e1000557.

44. Bjorkman, P., et al., *The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens.* Nature, 1987. 329(6139): p. 512.

45. Nei, M. and T. Gojobori, *Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.* Mol Biol Evol, 1986. 3(5): p. 418-26.

46. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.* Molecular biology and evolution, 2011. 28(10): p. 2731-2739.

47. Yang, Z., *PAML 4: phylogenetic analysis by maximum likelihood.* Molecular biology and evolution, 2007. 24(8): p. 1586-1591.

48. Pond, S.L.K. and S.D. Frost, *Datamonkey: rapid detection of selective pressure on individual sites of codon alignments.*Bioinformatics, 2005. 21(10): p. 2531-2533.

49. Murrell, B., et al., *Detecting individual sites subject to episodic diversifying selection.* PLoS genetics, 2012. 8(7): p. e1002764.

50. Kosakovsky Pond, S.L. and S.D. Frost, *Not so different after all: a comparison of methods for detecting amino acid sites under selection.* Molecular biology and evolution, 2005. 22(5): p. 1208-1222.

51. Murrell, B., et al., *FUBAR: a fast, unconstrained bayesian approximation for inferring selection.* Molecular biology and evolution, 2013. 30(5): p. 1196-1205.

52. Lecocq, T., et al., *Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (Bombus lapidarius).* BMC evolutionary biology, 2013. 13(1): p. 263.

53. Bozdogan, H., *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions.*Psychometrika, 1987. 52(3): p. 345-370.

54. Ronquist, F. and J.P. Huelsenbeck, *MrBayes 3: Bayesian phylogenetic inference under mixed models.* Bioinformatics, 2003. 19(12): p. 1572-1574.

55. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.*Molecular biology and evolution, 2016. 33(7): p. 1870-1874.

56. Minias, P., et al., *A global analysis of selection at the avian MHC.*Evolution, 2018. 72(6): p. 1278-1293.

57. Bos, D.H. and B. Waldman, *Evolution by recombination and transspecies polymorphism in the MHC class I gene of Xenopus laevis.*Molecular Biology and Evolution, 2005. 23(1): p. 137-143.

58. Loiseau, C., et al., *Diversifying selection on MHC class I in the house sparrow (Passer domesticus).* Molecular Ecology, 2009. 18(7): p. 1331-1340.

59. Freeman-Gallant, C.R., et al., *Variation at the major histocompatibility complex in Savannah sparrows.* Molecular Ecology, 2002. 11(6): p. 1125-1130.

60. Alcaide, M., M. Liu, and S.V. Edwards, *Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3.* PeerJ, 2013. 1: p. e86.

61. Nei, M. and A.P. Rooney, *Concerted and birth-and-death evolution of multigene families.* Annu. Rev. Genet., 2005. 39: p. 121-152.

62. Miller, H.C. and D.M. Lambert, *Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae).*Immunogenetics, 2004. 56(3): p. 178-191.

63. Burri, R., et al., *Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution.* Molecular Biology and Evolution, 2008. 25(6): p. 1180-1191.

64. Kikkawa, E.F., et al., *Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus).*Immunogenetics, 2009. 61(5): p. 341-352.

65. Eimes, J.A., et al., *Early duplication of a single MHC IIB locus prior to the passerine radiations.* PloS one, 2016. 11(9): p. e0163456.

66. Hughes, A.L., T. Ota, and M. Nei, *Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules.* Molecular biology and evolution, 1990. 7(6): p. 515-524.

67. Penn, D.J., K. Damjanovich, and W.K. Potts, *MHC heterozygosity confers a selective advantage against multiple-strain infections.*Proceedings of the National Academy of Sciences, 2002. 99(17): p. 11260-11264.

68. Bollmer, J.L., F.H. Vargas, and P.G. Parker, *Low MHC variation in the endangered Galapagos penguin (Spheniscus mendiculus).*Immunogenetics, 2007. 59(7): p. 593-602.

69. Sepil, I., S. Lachish, and B.C. Sheldon, *Mhc-linked survival and lifetime reproductive success in a wild population of great tits.*Molecular ecology, 2013. 22(2): p. 384-396.

70. Schaschl, H., et al., *Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (Rupicapra spp.).*Immunogenetics, 2005. 57(1-2): p. 108-115.

71. Minias, P., et al., *Contrasting evolutionary histories of MHC class I and class II loci in grouse—effects of selection and gene conversion.* Heredity, 2016. 116(5): p. 466.

72. Anmarkrud, J., et al., *Ancestral polymorphism in exon 2 of bluethroat (Luscinia svecica) MHC class II B genes.* Journal of Evolutionary Biology, 2010. 23(6): p. 1206-1217.

73. Zeng, Q.-Q., et al., *Balancing selection and recombination as evolutionary forces caused population genetic variations in golden pheasant MHC class I genes.* BMC evolutionary biology, 2016. 16(1): p. 42.

74. Wynne, J.W., et al., *Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (Salmo salar L.).* Fish & shellfish immunology, 2007. 22(6): p. 707-717.

75. Borg, A.A., et al., *Variation in MHC genotypes in two populations of house sparrow (Passer domesticus) with different population histories.* Ecology and evolution, 2011. 1(2): p. 145-159.

76. Ye, Q., et al., *Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant.* PloS one, 2012. 7(3): p. e32154.

77. Gillingham, M.A., et al., *Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species.* Journal of evolutionary biology, 2016. 29(2): p. 438-454.

78. Marmesat, E., et al., *Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus.*BMC evolutionary biology, 2017. 17(1): p. 158.

79. Jaratlerdsiri, W., et al., *Selection and trans-species polymorphism of major histocompatibility complex class II genes in the order Crocodylia.* PLoS One, 2014. 9(2): p. e87534.

80. Ballingall, K.T., et al., *Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep.* PLoS One, 2010. 5(6): p. e11402.