Research Article

Molecular Characterization of MHC Class I Genes in Four Species of the *Turdidae* Family to Assess Genetic Diversity and Selection

Muhammad Usman Ghani [iD],[1] **Li Bo** [iD],[1] **An Buyang** [iD],[2] **Xu Yanchun** [iD],[1] **Shakeel Hussain,**[1] **and Muhammad Yasir** [iD][3]

[1]*College of Wildlife Resources and Protected Area, Northeast Forestry University, Harbin 150040, China*
[2]*Department of Stem Cell Biology and Medicine, Graduate School of Medical Science, Kyushu University, Fukuoka 810-0000, Japan*
[3]*Department of Life Science and Technology, Huazhong Agricultural University, Wuhan, China*

Correspondence should be addressed to Muhammad Usman Ghani; drusmanghani466@gmail.com
and Li Bo; libo_770206@126.com

In vertebrate animals, the molecules encoded by major histocompatibility complex (MHC) genes play an essential role in the adaptive immunity. MHC class I deals 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 the *Turdidae* family from a broad geographical area of northeast China. We isolated 77 MHC class I sequences, including 47 putatively functional sequences and 30 pseudosequences from 80 individuals. Using the method based on analysis of cloned amplicons ($n = 25$) for each species, we found two and seven MHC I sequences per individual indicating more than one MHC I locus 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 are found to be divergent with overall polymorphic sites per species ranging between 58 and 70 (out of 291 sites). Moreover, transspecies 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 the *Turdidae* family rather than between other passerines. Pronounced MHC I gene diversity is essential for the 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.

1. Introduction

The major histocompatibility complex (MHC) is a group of 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, are useful to cell surface glycoproteins that regulate the immune response. MHC class II molecules are heterodimers consisting of an 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]. Heterodimer molecules of MHC class I are made up of an α chain and a non-MHC molecule, the $\beta2$ microglobulin. The α chain constitutes a cytoplasmic tail, a transmembrane domain, and three extracellular domains named $\alpha1$, $\alpha2$, and $\alpha3$ [3] that are encoded by exons 2, 3, and 4, respectively. 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 MHC class I [5]. Maintenance of

surprising diversity is supposed to 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 a 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 the pathogen community with time and locality results in MHC variation in the host population. Generally, in an individual possessing huge numbers and diverse MHC alleles; more pathogens can be recognized [1].

Structural diversity and immune response have been explored in numerous research, including genomics [7, 8], ailment [9–11], and 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 a 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 the 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) Measured locus-specific variation in MHC I exon 3 genes across the *Turdidae* family to evaluate 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 the *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 the *Turdidae* family and other avian species.

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 geographical location of sample material is presented (Figure 1).

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 (10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1% SDS). Total genomic DNA was extracted with AxyPrep Multisource Genomic DNA Miniprep Kit (AXYGEN, China) according to the manufacturer's instructions. The DNA concentration was measured with Nanopore Spectrophotometer at 260 nm absorbance. Samples above 100 ng/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$'$-TCCCCACAGGTCTC CACACAGT-3$'$ and HN46 reverse 5$'$-ATCCCAAATTC CCACCCACCTT-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. Amplification was performed in the reaction mixture containing 20 ng DNA template, 0.2 μM of each primer, 25 μl 2× EasyTaq® PCR SuperMix (+dye) (Trans, China), and water (deionized) to reach 50 μl 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™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
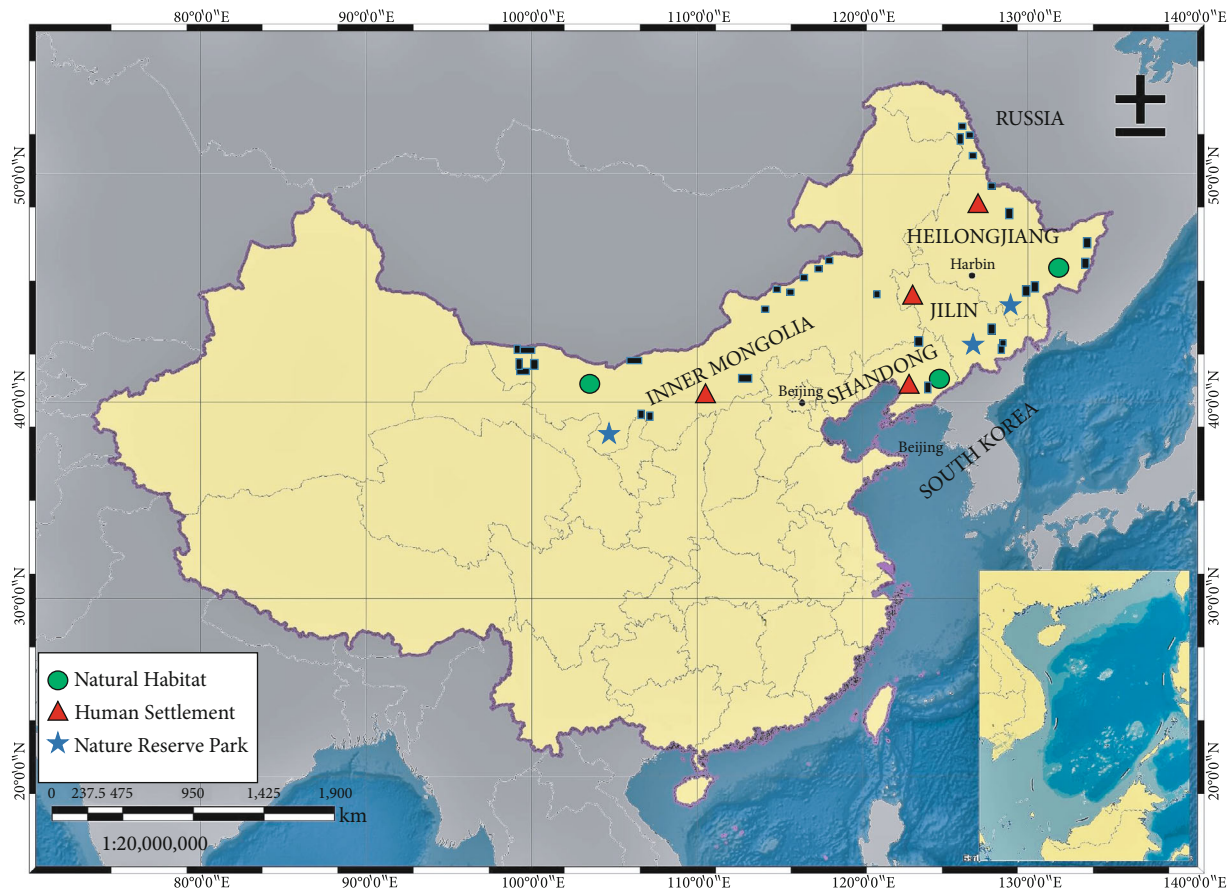
FIGURE 1: Geographical locations of samples included in our study. Square represents the actual site of the sample, and size of the square represents the approximate diameter of the sample's geographical range.

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 DNAStar7.1 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 nonsense 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].

2.6. 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 [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].

2.7. 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 and Li's $F*$, and Fu and 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 (2ΔlnL) 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].

2.8. 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*, and *Sturnidae*. 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 the *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

TABLE 1: Amplification success and genetic diversity within each of the four species of the *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 naumanni	285	64	33.7	0.118
				0.211
				0.091
				0.113
Turdus eunomus	285	58	32.32	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

dataset included 285-291 bp (Primers not include). Analysis of gDNA alignment revealed a total of 77 distinct Haplotypes/alleles including 47 PFA. 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 the *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 and five in most individuals of study population, and six of the thirteen pseudogene

TABLE 2: Recombinants detected in *Turdidae* family MHC class I alleles, parent sequences and 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*	Unknown (*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	

Note: NS indicates not significant. Maj *P* and Min *P* represent major and minor parents, respectively. BP denotes 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 those BPs. ** indicates $P < 0.01$.

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 eunomus* (0.113). The average number of nucleotides difference (*K*) varied between 43.95 in *Turdus atrogularis* and 32.32 in *Turdus eunomus*.

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 the *Turdidae* family. We only find one potential recombination event in *Tuna-MHCI * PFA06* in *Turdus naumanni* 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.

3.4. 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 eunomus* (Tajima's *D*: -0.87309, $P > 0.10$; Fu and Li's $D*$ test statistic: 0.36, $P > 0.10$; Fu and Li's $F*$ test statistic: 0.03, $P > 0.10$) and *Turdus atrogularis* (Tajima's *D*: -0.86107, $P > 0.10$ Fu and Li's $D*$ test statistic: 0.19, $P > 0.10$; Fu and Li's $F*$ test statistic: -0.077, $P > 0.10$). Still, overall d_N value was significantly higher statistically than d_S in *Turdus atrogularis* (1.687) and ratio d_N/d_S was more pronounced at codons presumably coding PBR (1.994) than codons not involved in such activity (0.884) is presented in (Table 3).

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).

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) and (Figure 3).

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

Species	Domain	$d_N \pm$ SE	$d_S \pm$ SE	Z	P	ω	$d_{nt} \pm$ SE	$d_{aa} \pm$ SE
	All sites	0.142 ± 0.026	0.104 ± 0.013	0.610	0.543	1.365	0.051 ± 0.08	0.246 ± 0.035
Turdus naumanni	PBR	0.281 ± 0.059	0.153 ± 0.032	1.848	**0.034**	1.835	0.041 ± 0.045	0.372 ± 0.082
	Non-PBR	0.051 ± 0.062	0.067 ± 0.012	1.356	0.476	0.761	0.034 ± 0.09	0.126 ± 0.029
	All sites	0.134 ± 0.021	0.107 ± 0.019	0.729	0.457	1.251	0.043 ± 0.010	0.202 ± 0.031
Turdus eunomus	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
	All sites	0.146 ± 0.048	0.101 ± 0.025	0.911	0.771	1.445	0.059 ± 0.017	0.191 ± 0.035
Turdus ruficollis	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
	All sites	0.189 ± 0.091	0.112 ± 0.025	1.040	0.150	1.687	0.063 ± 0.053	0.211 ± 0.039
Turdus atrogularis	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 which are in parentheses. Bold represents significant results.

TABLE 4: Estimation of d_N and d_S substitution rates for sites positively selected and their ratio for codons chosen a priori (PBR and non-PBR).

Species	Comparison	Model	lnL value	Parameter estimates	PSSs	LRT	TS value
Turdus naumanni	Tuna 1-30	M7 (beta)	-703.53	$P = 0.19$, $q = 0.138$, $P_0 = 0.351$	Not allowed	M7 vs. M8	5.47
		M8 (beta and omega)	-722.49	$P = 0.87$, $q = 0.117$, $P_1 = 0.09$, $\omega = 3.11$	39F,41L 88T		
Turdus eunomus	Tuna 1-30	M7 (beta)	-692.13	$P = 0.13$, $q = 0.97$, $P_0 = 0.347$	Not allowed	M7 vs. M8	4.56
		M8 (beta and omega)	-714.11	$P = 0.91$, $q = 0.158$, $P_1 = 0.11$, $\omega = 3.76$	41L, 52P,88T		
Turdus ruficollis	Turu1-15	M7 (beta)	-811.27	$P = 0.49$, $q = 0.187$, $P_0 = 0.411$	Not allowed	M7 vs. M8	6.21
		M8 (beta and omega)	-834.88	$P = 0.131$, $q = 0.232$, $P_1 = 0.13$, $\omega = 3.94$	29Y, **78F**, 88G		
Turdus atrogularis	Tuat1-5	M7 (beta)	-847.53	$P = 0.83$, $q = 0.211$, $P_0 = 0.585$	Not allowed	M7 vs. M8	7.10
		M8 (beta and omega)	-849.78	$P = 0.173$, $q = 0.279$, $P_1 = 0.17$, $\omega = 4.11$	39H, 78F, **91Q**		

*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%.

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.

3.5. Phylogenetic Analysis.

In phylogenetic analysis, we observed that sampled species form a well-supported monophyletic clade with *Erithacus rubeculs* members of the *Turdidae* family in maximum likelihood analysis. 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
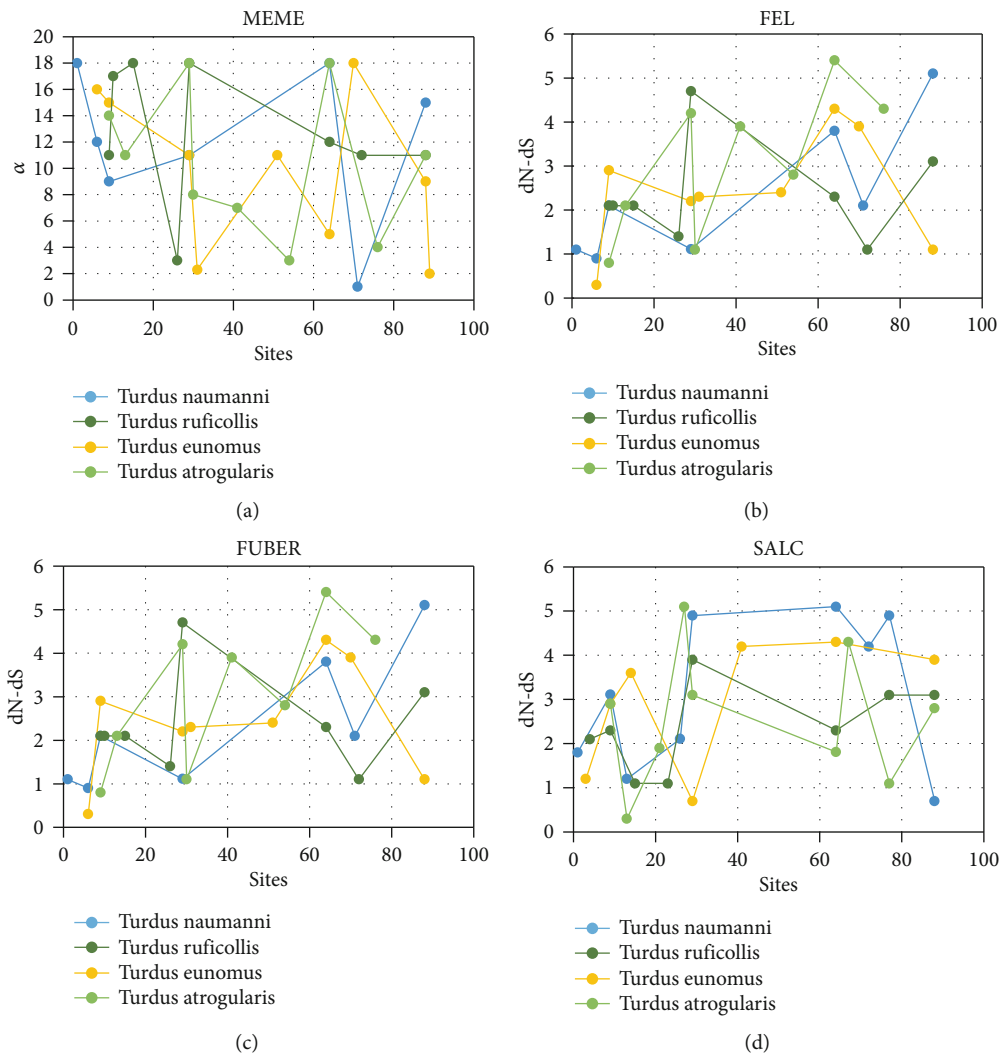
FIGURE 2: Positively selected sites using the online adoptive evolutionary server Datamonkey with (a) MEME, (b) FEL, (c) FUBAR, and (d) SALC. Substitution tests identify numerous codons showing the signature of positive selection.

the *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 transspecies polymorphism or orthology [57].

4. Discussion

In this study, we have for the first time characterize MHC Class I gene in four species of the *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 rufficollis* 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
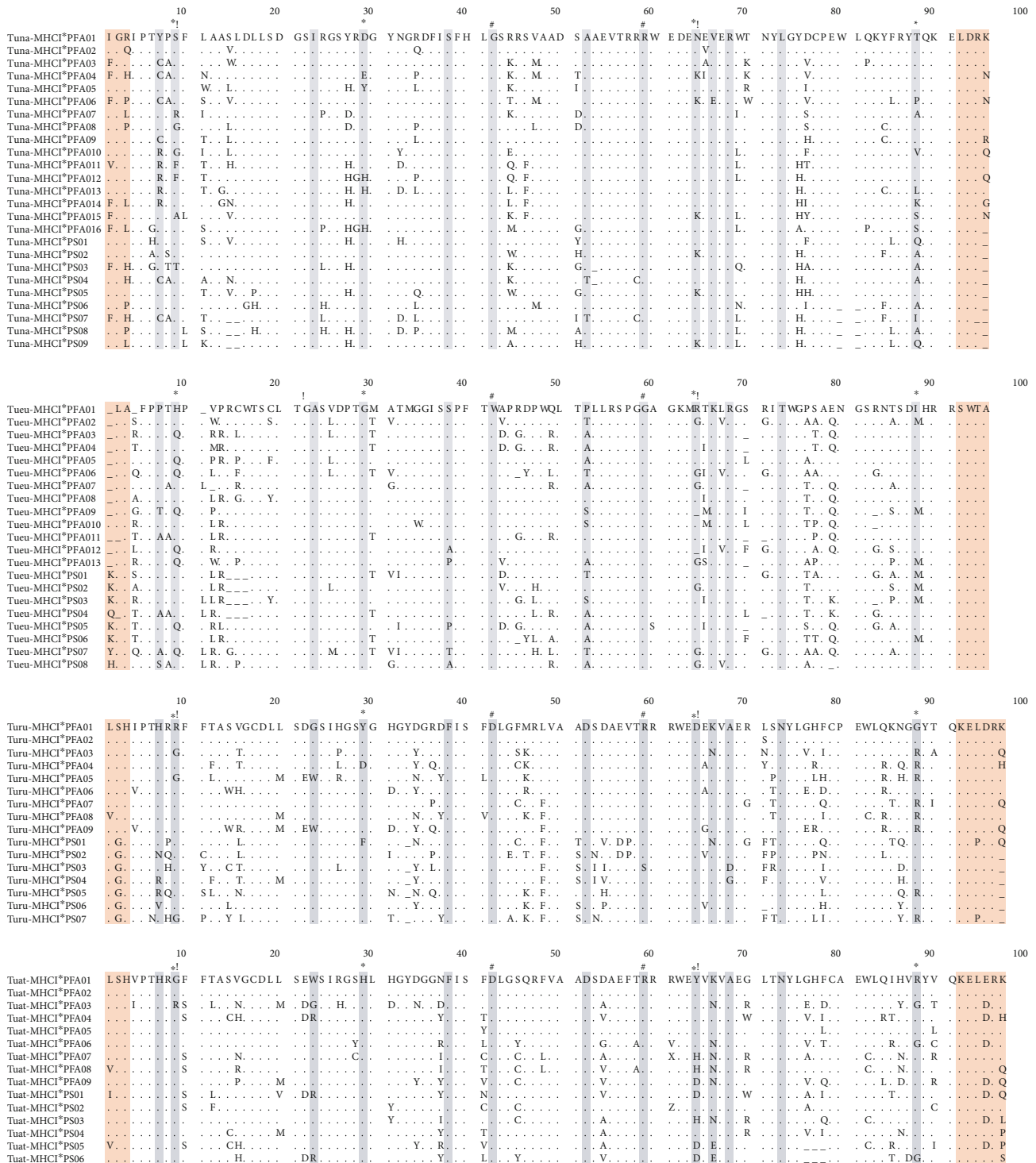
Tuna-MHCI*PFA01 I GRI PTYPS F L AAS LDLLS D GSI RGS YRDG YNGRDFI S FH L GS RRS VAAD S AAE VTRRRW E DENE VERWT NYLGYDCPEW L QKYF RYTQK E L DRK

Tueu-MHCI*PFA01 _ L A _ F P PTHP _ VP RCWTS CL T GAS VDPT GM A TMGGI S S PF T WAP RDP WQL T PLLRS P GGA GKMRTKLRGS R I TWGP S AEN GS RNTS DI HR R S WT A

Turu-MHCI*PFA01 L SH I PTHRRF FTAS VGCDLL SDGS I HGS YG HGYDGRDFI S FDLGFMRLVA ADS DAEVTRR RWEDEK VAER L SNYLGHFCP EWLQKNGGYT QKE L DRK

Tuat-MHCI*PFA01 L SH VPTHRGF FTAS VGCDLL S EWS I RGSHL HGYDGGNFI S FDLGSQRFVA ADS DAEFTRR RWE YVK VAEG L TNYLGHFCA EWLQI HVRYV QKE L ERK

Figure 3: Alignment of deduced amino acid sequences of all alleles retrieved from four species of the *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. The light brown region indicates flanking introns. - indicates missing nucleotides. ∗ represents codons positively selected in almost all of the tests performed for selection analysis. ! indicates the homologue region also positively selected in most of avian species. # inferred homology with human.

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 func-

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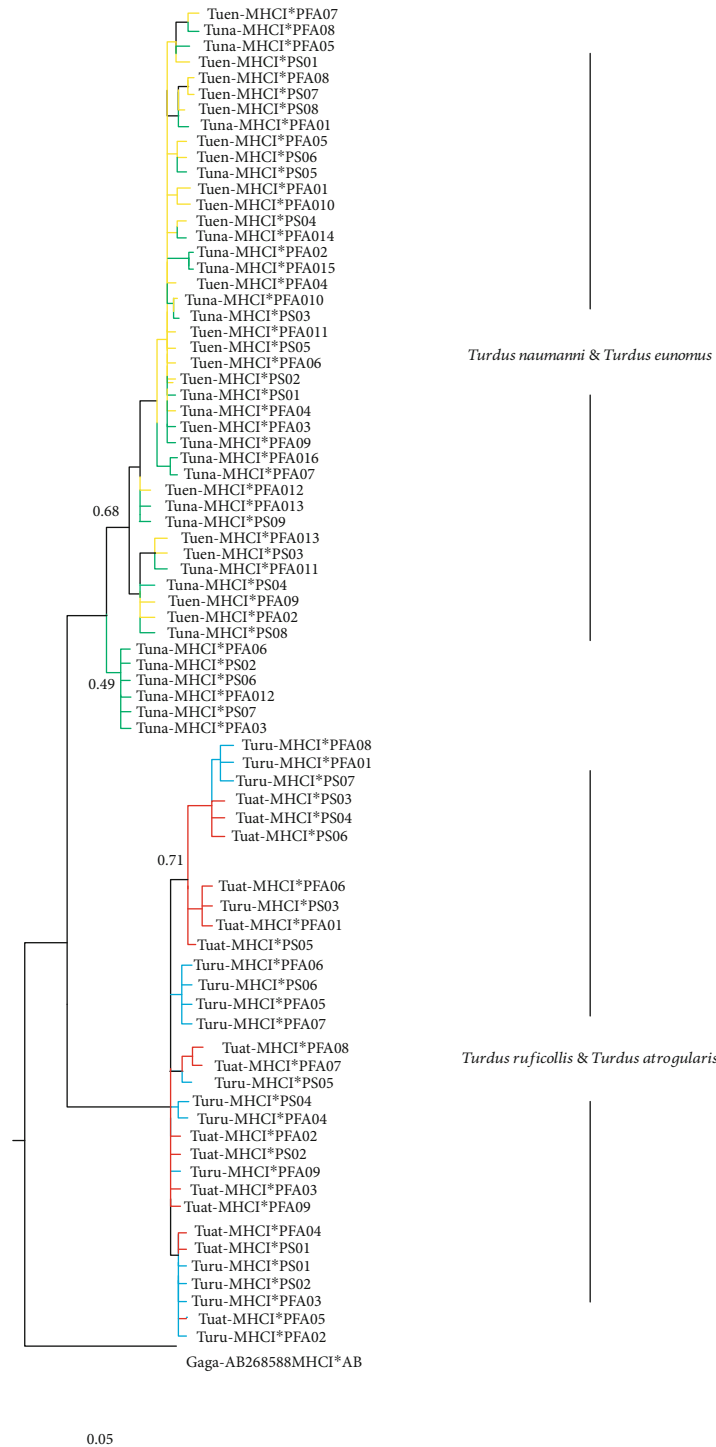
FIGURE 4: Bayesian phylogenetic reconstruction of MHC class I exon 3 of four species of the *Turdidae* family. All the nodes are well supported (PP > 0.90%) unless indicated otherwise. AB268885 *Gallus gallus* was used as an outgroup.

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 recom-

bination 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
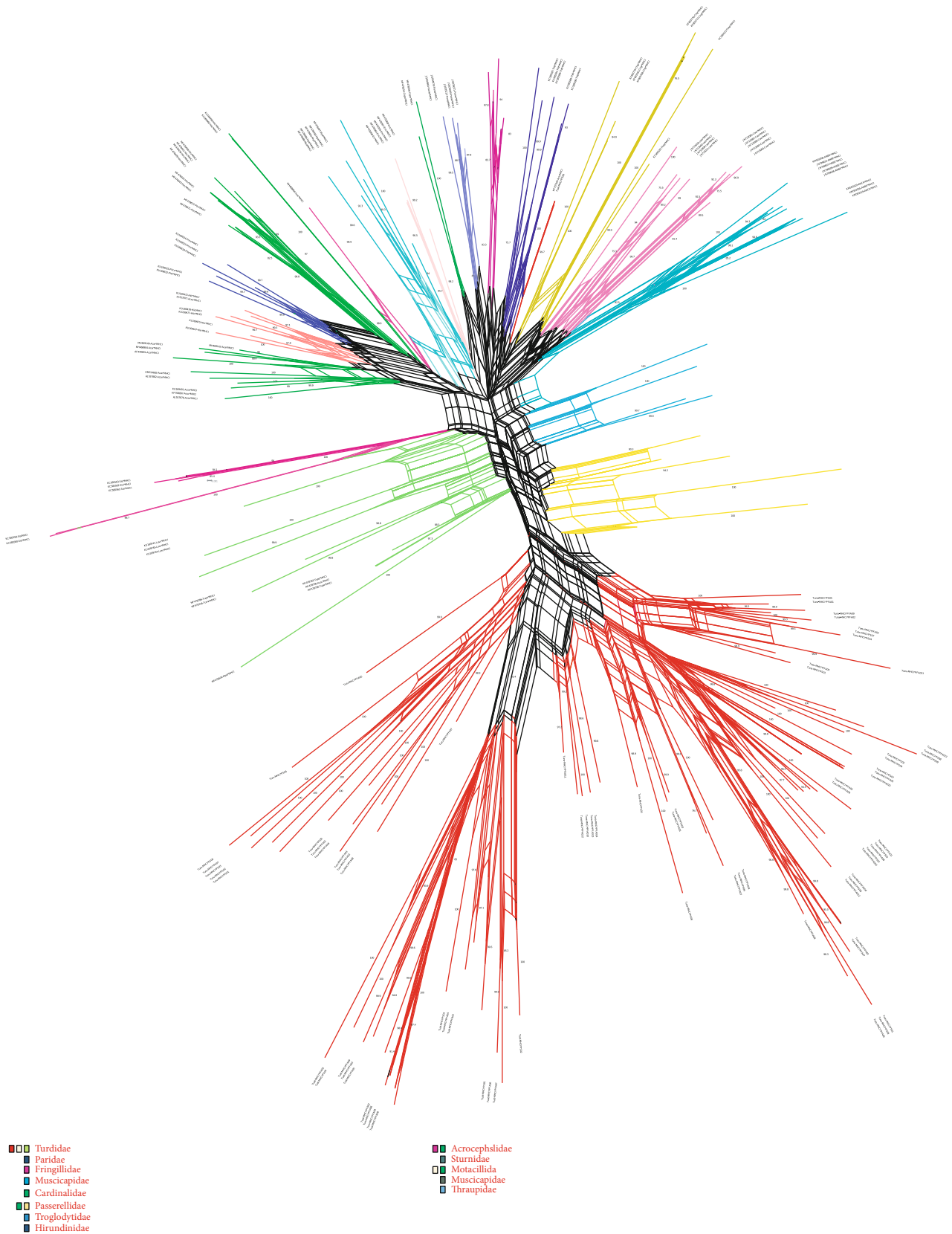
FIGURE 5: The phylogenetic networks of MHC class I of four species of the *Turdidae* family along with homologue sequences from passerine species. GenBank accession numbers are provided. Species names are mentioned in the lower right side with branch colors. Neighbor-Net networks based on uncorrected *P*-distances and carried out in 1000 bootstrap replicates to estimate nodal support. Nodal support values (>75%) were displayed.

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 and Li's $D*$ and Fu and 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 $\alpha 2$ 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 the *Turdidae* family clustered together with sequences from congeneric species. We found increased sequences similarities between the 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).

5. Conclusion

Our study shows that species of the *Turdidae* family has retained significant MHC class I diversity, which supports high conservational value and contributes to the evolution of MHC class I genes. Importantly, we specifically amplify the exon 3 locus and 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 which supported the hypothesis that different mechanisms can shape evolutionary paths of MHC class I.

Abbreviations

MHC: Major histocompatibility complex
PBR: Peptide-binding region
CDs: Cluster of differentiation
CDRs: Complementary-determining regions
ABS: Antigen-binding site
TCR: T-cell receptors
SDS: Sodium dodocyl sulfate
EDTA: Ethylene diamine tetra acetic acid
PCR: Polymerase chain reaction
HLA: Human leukocyte antigen
MEGA: Molecular Evolutionary Genetics Analysis
GTR: General time-reversible model
PFA: Putative function alleles
GARD: Genetic algorithm for recombination detection
RDP: Recombination detection program
PP: Posterior probability
BP: Breakpoint
d_N: Nonsynonymous substitution
d_S: Synonymous substitution
d_{nt}: Nucleotide distance
d_{aa}: Amino acid distance
SE: Standard error
PSSs: Positively selected sites
BEB: Bayes empirical Bayes
MEME: Mixed effects model of evolution
SALC: Single likelihood ancestor counting FEL, fixed effect likelihood
FUBAR: Fast unconstrained Bayesian approximation
PAML: Phylogenetic analysis using maximum likelihood
TSP: Transspecies polymorphism.

Data Availability

The data of this study will be available openly to readers, and they can access the data supporting the conclusions of the study.

Disclosure

The manuscript has been presented in "pre-print" at https://www.researchgate.net/publication/346148804.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

MUG, AY, and LB designed the study. MUG carried out the experiment and drafted the experiment. LB supervised the whole study, provided recommendations for, and revised the MS.YCX provided valuable suggestion for the MS. All authors contributed to and approved the current manuscript draft.

Acknowledgments

References

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

[2] K. K. Jensen, M. Andreatta, P. Marcatili et al., "Improved methods for predicting peptide binding affinity to MHC class II molecules," *Immunology*, vol. 154, no. 3, pp. 394–406, 2018.

[3] P. J. Bjorkman and P. Parham, "Structure, function, and diversity of class I major histocompatibility complex molecules," *Annual Review of Biochemistry*, vol. 59, no. 1, pp. 253–288, 1990.

[4] X.-H. Zhang, Z. X. Dai, G. H. Zhang, J. B. Han, and Y. T. Zheng, "Molecular characterization, balancing selection, and genomic organization of the tree shrew (*Tupaia belangeri*) MHC class I gene," *Gene*, vol. 522, no. 2, pp. 147–155, 2013.

[5] M. Yeager and A. L. Hughes, "Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution," *Immunological Reviews*, vol. 167, no. 1, pp. 45–58, 1999.

[6] S. Piertney and M. Oliver, "The evolutionary ecology of the major histocompatibility complex," *Heredity*, vol. 96, no. 1, pp. 7–21, 2006.

[7] J. K. Kulski, T. Shiina, T. Anzai, S. Kohara, and H. Inoko, "Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man," *Immunological Reviews*, vol. 190, no. 1, pp. 95–122, 2002.

[8] J. Kelley, L. Walter, and J. Trowsdale, "Comparative genomics of major histocompatibility complexes," *Immunogenetics*, vol. 56, no. 10, pp. 683–695, 2005.

[9] Å. Langefors, J. Lohm, M. Grahn, Ø. Andersen, and T. . Schantz, "Association between major histocompatibility complex class IIB alleles and resistance toAeromonas salmonicidain Atlantic salmon," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1466, pp. 479–485, 2001.

[10] P. W. Hedrick, "Pathogen resistance and genetic variation at MHC loci," *Evolution*, vol. 56, no. 10, pp. 1902–1908, 2002.

[11] H. Westerdahl, "Passerine MHC: genetic variation and disease resistance in the wild," *Journal of Ornithology*, vol. 148, no. S2, pp. 469–477, 2007.

[12] S. Paterson 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*, vol. 264, no. 1389, pp. 1813–1819, 1997.

[13] C. Landry, D. Garant, P. Duchesne, and L. Bernatchez, "'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*, vol. 268, no. 1473, pp. 1279–1285, 2001.

[14] G. J. Knafler, J. A. Clark, P. D. Boersma, and J. L. Bouzat, "MHC diversity and mate choice in the Magellanic penguin, Spheniscus magellanicus," *Journal of Heredity*, vol. 103, no. 6, pp. 759–768, 2012.

[15] J. A. Borghans, J. B. Beltman, and R. J. De Boer, "MHC polymorphism under host-pathogen coevolution," *Immunogenetics*, vol. 55, no. 11, pp. 732–739, 2004.

[16] A. Jepson, W. Banya, F. Sisay-Joof 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*, vol. 65, no. 3, pp. 872–876, 1997.

[17] S. V. Edwards, J. Gasper, D. Garrigan, D. Martindale, and B. F. Koop, "A 39-kb sequence around a blackbird Mhc class II gene: ghost of selection past and songbird genome architecture," *Molecular Biology and Evolution*, vol. 17, no. 9, pp. 1384–1395, 2000.

[18] D. Canal, M. Alcaide, J. A. Anmarkrud, and J. Potti, "Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher Ficedula hypoleuca," *BMC Research Notes*, vol. 3, no. 1, p. 236, 2010.

[19] G. Kroemer, A. Bernot, G. Behar et al., "Molecular genetics of the chicken MHC: current status and evolutionary aspects," *Immunological Reviews*, vol. 113, no. 1, pp. 119–145, 1990.

[20] R. Zoorob, A. Bernot, D. M. Renoir, F. Choukri, and C. Auffray, "Chicken major histocompatibility complex class II B genes: analysis of interallelic and inter-locus sequence variance," *European Journal of Immunology*, vol. 23, no. 5, pp. 1139–1145, 1993.

[21] M. M. Peacock and A. T. Smith, "The effect of habitat fragmentation on dispersal patterns, mating behavior, and genetic variation in a pika (*Ochotona princeps*) metapopulation," *Oecologia*, vol. 112, no. 4, pp. 524–533, 1997.

[22] C. Bonneaud, G. Sorci, V. Morin, H. Westerdahl, R. Zoorob, and H. Wittzell, "Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*)," *Immunogenetics*, vol. 55, no. 12, pp. 855–865, 2004.

[23] H. Westerdahl, "No evidence of an MHC-based female mating preference in great reed warblers," *Molecular Ecology*, vol. 13, no. 8, pp. 2465–2470, 2004.

[24] M. Promerová, 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*, vol. 61, no. 6, pp. 451–461, 2009.

[25] W. BABIK, P. TABERLET, M. J. EJSMOND, and J. RADWAN, "New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system," *Molecular Ecology Resources*, vol. 9, no. 3, pp. 713–719, 2009.

[26] T. Kanagawa, "Bias and artifacts in multitemplate polymerase chain reactions (PCR)," *Journal of Bioscience and Bioengineering*, vol. 96, no. 4, pp. 317–323, 2003.

[27] S. Abduriyim, Y. Nishita, P. A. Kosintsev et al., "*Evolution* of MHC class I genes in Eurasian badgers, genus *Meles* (Carnivora, Mustelidae)," *Heredity*, vol. 122, no. 2, pp. 205–218, 2019.

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

[29] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947-2948, 2007.

[30] J. Klein, R. E. Bontrop, R. L. Dawkins et al., *Nomenclature for the major histocompatibility complexes of different species: a proposal*, The HLA system in clinical transplantation, Springer, 1993.

[31] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[32] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451-1452, 2009.

[33] D. Martin, D. Posada, K. A. Crandall, and C. Williamson, "A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints," *AIDS Research & Human Retroviruses*, vol. 21, no. 1, pp. 98–102, 2005.

[34] M. Padidam, S. Sawyer, and C. M. Fauquet, "Possible emergence of new geminiviruses by frequent recombination," *Virology*, vol. 265, no. 2, pp. 218–225, 1999.

[35] D. Posada, "jModelTest: phylogenetic model averaging," *Molecular Biology and Evolution*, vol. 25, no. 7, pp. 1253–1256, 2008.

[36] J. M. Smith, "Analyzing the mosaic structure of genes," *Journal of Molecular Evolution*, vol. 34, no. 2, pp. 126–129, 1992.

[37] M. J. Gibbs, J. S. Armstrong, and A. J. Gibbs, "Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences," *Bioinformatics*, vol. 16, no. 7, pp. 573–582, 2000.

[38] M. F. Boni, D. Posada, and M. W. Feldman, "An exact nonparametric method for inferring mosaic structure in sequence triplets," *Genetics*, vol. 176, no. 2, pp. 1035–1047, 2007.

[39] S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost, "GARD: a genetic algorithm for recombination detection," *Bioinformatics*, vol. 22, no. 24, pp. 3096–3098, 2006.

[40] C. N. Balakrishnan, R. Ekblom, M. Völker et al., "Gene duplication and fragmentation in the zebra finch major histocompatibility complex," *BMC Biology*, vol. 8, no. 1, p. 29, 2010.

[41] M. Alcaide, J. Muñoz, J. Martínez-de la Puente, R. Soriguer, and J. Figuerola, "Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian coot Fulica atra (Aves: Rallidae)," *Ecology and Evolution*, vol. 4, no. 6, pp. 688–698, 2014.

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

[43] C. S. Hee, S. Gao, B. Loll et al., "Structure of a classical MHC class I molecule that binds "non-classical" ligands," *PLoS Biology*, vol. 8, no. 12, article e1000557, 2010.

[44] P. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley, "The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens," *Nature*, vol. 329, no. 6139, pp. 512–518, 1987.

[45] M. Nei and T. Gojobori, "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions," *Molecular Biology and Evolution*, vol. 3, no. 5, pp. 418–426, 1986.

[46] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[47] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.

[48] S. L. K. Pond and S. D. Frost, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments," *Bioinformatics*, vol. 21, no. 10, pp. 2531–2533, 2005.

[49] B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. Kosakovsky Pond, "Detecting individual sites subject to episodic diversifying selection," *PLoS Genetics*, vol. 8, no. 7, article e1002764, 2012.

[50] S. L. Kosakovsky Pond and S. D. Frost, "Not so different after all: a comparison of methods for detecting amino acid sites under selection," *Molecular Biology and Evolution*, vol. 22, no. 5, pp. 1208–1222, 2005.

[51] B. Murrell, S. Moola, A. Mabona et al., "FUBAR: a fast, unconstrained Bayesian approximation for inferring selection," *Molecular Biology and Evolution*, vol. 30, no. 5, pp. 1196–1205, 2013.

[52] T. Lecocq, S. Dellicour, D. Michez et al., "Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*)," *BMC Evolutionary Biology*, vol. 13, no. 1, p. 263, 2013.

[53] H. Bozdogan, "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.

[54] F. Ronquist and J. P. Huelsenbeck, "MrBayes 3: Bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, 2003.

[55] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[56] P. Minias, E. Pikus, L. A. Whittingham, and P. O. Dunn, "A global analysis of selection at the avian MHC," *Evolution*, vol. 72, no. 6, pp. 1278–1293, 2018.

[57] D. H. Bos and B. Waldman, "Evolution by recombination and transspecies polymorphism in the MHC class I gene of *Xenopus laevis*," *Molecular Biology and Evolution*, vol. 23, no. 1, pp. 137–143, 2006.

[58] C. Loiseau, M. Richard, S. Garnier et al., "Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*)," *Molecular Ecology*, vol. 18, no. 7, pp. 1331–1340, 2009.

[59] C. R. Freeman-Gallant, E. M. Johnson, F. Saponara, and M. Stanger, "Variation at the major histocompatibility complex in Savannah sparrows," *Molecular Ecology*, vol. 11, no. 6, pp. 1125–1130, 2002.

[60] M. Alcaide, 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*, vol. 1, article e86, 2013.

[61] M. Nei and A. P. Rooney, "Concerted and birth-and-death evolution of multigene families," *Annual Review of Genetics*, vol. 39, no. 1, pp. 121–152, 2005.

[62] H. C. Miller and D. M. Lambert, "Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae)," *Immunogenetics*, vol. 56, no. 3, pp. 178–191, 2004.

[63] R. Burri, H. N. Hirzel, N. Salamin, A. Roulin, and L. Fumagalli, "Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution," *Molecular Biology and Evolution*, vol. 25, no. 6, pp. 1180–1191, 2008.

[64] E. F. Kikkawa, T. T. Tsuda, D. Sumiyama et al., "Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*)," *Immunogenetics*, vol. 61, no. 5, pp. 341–352, 2009.

[65] J. A. Eimes, S. I. Lee, A. K. Townsend, P. Jablonski, I. Nishiumi, and Y. Satta, "Early duplication of a single MHC IIB locus prior to the passerine radiations," *PLoS One*, vol. 11, no. 9, article e0163456, 2016.

[66] A. L. Hughes, 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*, vol. 7, no. 6, pp. 515–524, 1990.

[67] D. J. Penn, K. Damjanovich, and W. K. Potts, "MHC heterozygosity confers a selective advantage against multiple-strain infections," *Proceedings of the National Academy of Sciences*, vol. 99, no. 17, pp. 11260–11264, 2002.

[68] J. L. Bollmer, F. H. Vargas, and P. G. Parker, "Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*)," *Immunogenetics*, vol. 59, no. 7, pp. 593–602, 2007.

[69] I. Sepil, S. Lachish, and B. C. Sheldon, "Mhc-linked survival and lifetime reproductive success in a wild population of great tits," *Molecular Ecology*, vol. 22, no. 2, pp. 384–396, 2013.

[70] H. Schaschl, F. Suchentrunk, S. Hammer, and S. J. Goodman, "Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.)," *Immunogenetics*, vol. 57, no. 1-2, pp. 108–115, 2005.

[71] P. Minias, Z. W. Bateson, L. A. Whittingham, J. A. Johnson, S. Oyler-McCance, and P. O. Dunn, "Contrasting evolutionary histories of MHC class I and class II loci in grouse–effects of selection and gene conversion," *Heredity*, vol. 116, no. 5, pp. 466–476, 2016.

[72] J. A. Anmarkrud, A. Johnsen, L. Bachmann, and J. T. Lifjeld, "Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes," *Journal of Evolutionary Biology*, vol. 23, no. 6, pp. 1206–1217, 2010.

[73] Q.-Q. Zeng, K. He, D. D. Sun et al., "Balancing selection and recombination as evolutionary forces caused population genetic variations in golden pheasant MHC class I genes," *BMC Evolutionary Biology*, vol. 16, no. 1, p. 42, 2016.

[74] J. W. Wynne, M. T. Cook, B. F. Nowak, and N. G. Elliott, "Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (*Salmo salar* L.)," *Fish & Shellfish Immunology*, vol. 22, no. 6, pp. 707–717, 2007.

[75] Å. A. Borg, S. A. Pedersen, H. Jensen, and H. Westerdahl, "Variation in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with different population histories," *Ecology and Evolution*, vol. 1, no. 2, pp. 145–159, 2011.

[76] Q. Ye, K. He, S. Y. Wu, and Q. H. Wan, "Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant," *PLoS One*, vol. 7, no. 3, article e32154, 2012.

[77] M. A. Gillingham, A. Courtiol, M. Teixeira, M. Galan, A. Bechet, and F. Cezilly, "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*, vol. 29, no. 2, pp. 438–454, 2016.

[78] E. Marmesat, K. Schmidt, A. P. Saveljev, I. V. Seryodkin, and J. A. Godoy, "Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus," *BMC Evolutionary Biology*, vol. 17, no. 1, p. 158, 2017.

[79] W. Jaratlerdsiri, S. R. Isberg, D. P. Higgins, L. G. Miles, and J. Gongora, "Selection and trans-species polymorphism of major histocompatibility complex class II genes in the order Crocodylia," *PLoS One*, vol. 9, no. 2, article e87534, 2014.

[80] K. T. Ballingall, M. S. Rocchi, D. J. McKeever, and F. Wright, "Trans-species polymorphism and selection in the MHC class II DRA genes of domestic sheep," *PLoS One*, vol. 5, no. 6, article e11402, 2010.