diversity

MDPI

Article

Evaluation of Genetic Diversity and Parasite-Mediated Selection of MHC Class I Genes in *Emberiza godlewskii* (Passeriformes: Emberizidae)

Wei Huang [1,2,3,†], Xinyi Wang [1,†], Boye Liu [1,4], Tobias L. Lenz [2], Yangyang Peng [1] and Lu Dong [1,*] and Yanyun Zhang [1,*]

1 MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China
2 Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, 20146 Hamburg, Germany
3 Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH8 9YL, UK
4 Shaanxi Institute of Zoology, Xi'an 710033, China
* Correspondence: donglu@bnu.edu.cn (L.D.); zhangyy@bnu.edu.cn (Y.Z.)
† These authors contributed equally to this work.

Abstract: The major histocompatibility complex (MHC) is a multi-copy immune gene family in vertebrates. Its genes are highly variable and code for antigen-presenting molecules. Characterization of MHC genes in different species and investigating the mechanisms that shape MHC diversity is an important goal in understanding the evolution of biological diversity. Here we developed a next-generation sequencing (NGS) protocol to genotype the MHC class I genes of 326 Godlewski's buntings (*Emberiza godlewskii*) sampled in the Western mountain area of Beijing from 2014 to 2016. A total of 184 functional alleles were identified, including both non-classical and classical alleles, clustering into nine supertypes. Compared with other passerine birds, the number of MHC class I alleles per individual in Godlewski's buntings is high (mean 16.1 ± 3.3, median 16). In addition, we demonstrated signatures of historical and contemporary selection on MHC genes. Reflecting historical selection, ten amino acid sites in the antigen-binding domain showed signatures of balancing selection, eight of which exhibit high amino acid polymorphism. In terms of contemporary selection, we found that specific MHC supertypes were nominally associated with the infection of two malaria parasite lineages. These findings indicate the action of historical and possibly also contemporary balancing selection and suggest negative frequency-dependent or fluctuating selection as possible selection mechanisms.

Keywords: MHC; Godlewski's bunting; positive selection; supertypes; malaria parasites

Citation: Huang, W.; Wang, X.; Liu, B.; Lenz, T.L.; Peng, Y.; Dong, L.; Zhang, Y. Evaluation of genetic diversity and parasite-mediated selection of MHC class I genes in *Emberiza godlewskii* (Passeriformes: Emberizidae). *Diversity* **2022**, *14*, 925. https://doi.org/10.3390/d14110925

Academic Editor: Elmira Mohandesan

Received: 20 September 2022
Accepted: 25 October 2022
Published: 29 October 2022

1. Introduction

The immune system is critical in pathogen resistance as it provides different mechanisms to protect the host from infection. Therefore, the genetic diversity of immune genes is assumed to be strongly linked with the infection of pathogens and is maintained through coevolution between hosts and pathogens [1]. Among the immune genes that have been intensively studied in vertebrates, major histocompatibility complex (MHC) genes play a crucially important role in adaptive immunity [2–5]. MHC molecules encoded by MHC genes are capable of binding both self-derived and pathogen-derived peptides and presenting them to T cells to invoke an adaptive immune response. There are two main classes of MHC genes: Class I genes encode molecules that are responsible for binding intracellular peptides while class II genes encode molecules that are responsible for binding extracellular peptides. MHC genes are among the most variable gene families in vertebrates and the mechanisms driving diversity of MHC genes have received considerable attention [4,6,7].

Parasite-mediated selection is thought to be a critical driver for the polymorphism of MHC diversity. Three main hypotheses have been proposed to elucidate the association between pathogens and MHC genes [6]. (1) Heterozygote advantage: individuals which are heterozygous across the MHC can recognize a greater variety of pathogen antigens thus variations at MHC loci are selected and maintained [8]. (2) Rare-allele advantage: there is strong selection on pathogens to evade immune protection provided by the most common MHC alleles resulting in rare MHC alleles having a selective advantage [9]. However, the advantage disappears when the frequency of a protective MHC allele increases and this will maintain MHC variations by negative frequency-dependent selection [8,10]. (3) Fluctuating selection: a changing environment results in different parasites being abundant in space and time which generates directional selection in different subpopulations at different time points to maintain MHC diversity [11]. Although each of the three hypotheses can be supported by certain studies, the relative importance of these three hypotheses is still not clear [6,7]. Besides the paradigm of parasite-mediated selection, sexual selection is viewed as another important force to drive MHC diversity [12]. Mate choice based on MHC diversity or MHC compatibility has been reported in several studies [13,14] while other studies reported negative results [15]. Similar with parasite-mediated selection, it is hard to draw a clear conclusion about MHC-dependent selection from these mixed results and thus more case studies are needed to figure out the detailed selection mechanisms that shape MHC variations.

The number of MHC gene copies varies between different species. In avian species, chickens are believed to have "Minimum essential MHC" with only two class I and class II genes [16]. However, Passerine birds possess the most variable and complex MHC systems in birds [17–19]. For instance, a study of MHC class I genes across Passerida has found the number of MHC alleles per individual varies from 3–7 in bluethroat (*Luscinia svecica*) to 27–35 in willow warbler (*Phylloscopus trochilus*) [20]. A more recent study of sedge warbler (*Acrocephalus schoenobaenus*) reported the highest diversity of MHC Class I genes in passerine birds with a maximum of 65 alleles per individual [21]. However, despite the high diversity of MHC genes in passerine birds, some alleles have limited expression (non-classical alleles) while others are highly expressed (classical allele). Patterns of classical alleles and non-classical alleles have been identified in three sparrow species which indicates the complexity of Passerine MHC genes [22]. Because of the high diversity of MHC genes in passerine birds, it is crucially important to carefully characterize MHC genes in passerine birds in order to understand the mechanisms shaping MHC diversity. With the application of next generation sequencing (NGS) methods in MHC genotyping [23,24], passerine MHC genes have attracted considerable attention and the number of related studies has increased rapidly [21,25–29]. Avian malaria is widespread among passerine birds [30,31] and has been found to be associated with MHC genes in some passerine species [17], indicating the potential role of malaria parasites in shaping the diversity of MHC genes. However, it is not clear which selection mechanism is operating based on the evidence from previous studies that examined association between MHC variation and malaria parasites. Some studies identified associations between MHC diversity and malaria parasites [28,32] while others demonstrated associations between specific MHC alleles or supertypes and malaria parasites [29,33]. Thus, it is important to examine MHC-malaria association in a wide range of taxa to figure out the specific selection mechanism [28,33–36].

Godlewski's Bunting (*Emberiza godlewskii*) is a common resident bird in mountains of North China. Almost 90% of the Godlewski's Buntings in Beijing were infected by haemosporidian parasites and most of the infections were caused by three dominant *Plasmodium* lineages (ALARV04, PADOM02 and DENVID02) and one dominant *Leucocytozoon* lineage (EMSPO05) with annual stability [37], which makes it an interesting study system to explore the association between MHC variation and malaria parasites. In this study, we characterized MHC class I exon 3 of MHC class I genes in Godlewski's buntings. Exon 3 encodes an important part of the peptide binding domain of MHC class I molecules. MHC

class I genes are more polymorphic and associated with parasite resistance and mate choice in other passerine birds [34,38]. After establishing an NGS-based genotyping method to identify MHC alleles in Godlewski's buntings, our further analysis focused on the following aspects: (1) Characterize the population level and individual level of genetic diversity of MHC class I genes in Godlewski's buntings. (2) Determine the functional variation of MHC class I genes in Godlewski's buntings. (3) Identify signatures of historical and contemporary selection on MHC genes in Godlewski's buntings under malaria selection.

2. Materials and Methods

2.1. Sampling

We captured 326 male Godlewski's buntings by mist net and collected blood samples from brachial vein between 2014 and 2016 in Beijing, China (Sampling Approval by Beijing Normal University: No. CLS-EAW-2013-007) (Supplementary Materials Figure S1). Seven blood samples were stored in RNAfixer (Bioro yee, Beijing, China) for RNA extraction while the others were stored in ethanol.

2.2. DNA and RNA Extraction

Total genomic DNA was extracted from blood samples using the TIANamp genomic DNA kit (TIANGEN). RNA was extracted using the RNAstore Blood RNA Kit (Bioro yee) and reverse transcribed to complementary DNA (cDNA) using the PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Beijing, China) according to the manufacturer's protocol.

2.3. PCR Amplification and NGS Sequencing

To amplify the partial exon 3 of MHC class I genes, the primers (MHCD-F2: TWMYGGCTGTGAMCTYCY, MHCD-R2: TTGYRCTCCARCYCYTTC), refined from established primer pair MHCD-F/R, were used to amplify a fragment between 205 bp and 222 bp [27]. The MHCD-F/R was first selected after comparison with other pairs of degenerate primers (HN34/HN45 and MhcPasCI-FW/MhcPasCI-RV) and had the best performance in terms of efficiency and specificity [39,40]. As the fragments amplified by MHCD-F/R are included in those amplified by HN34/HN45, we compared the sequences generated by MHCD-F/R and HN34/HN45 and refined the primers based on the alignments such that some bases in MHCD-F/R were replaced with degenerate bases and vice versa. However, it is possible that the primers can only amplify a fraction of MHC class I alleles. Each of the two primers was extended with 9 bp tags at the 5′ end which consist of an identical 6 bp tag for library construction and a specific 3 bp tag to distinguish the amplicon from the pooling samples. PCR was performed in a 20 µL reaction with 1 µL of each primer, 1 µL DNA, 7 µL ddH_2O, and 10 µL PCR Mastermix (TaKaRa). PCR conditions included an initial denaturation at 94 degrees C for 5mins, followed by 35 cycles of denaturation at 94 degrees for 30 s, annealing at 63 degrees for 30 s and extension at 72 degrees for 30 s with a final step at 72 degrees for 7 min. The amplifications were performed in 96-well PCR plates with 6 negative controls per PCR plate. Successful PCR products verified by electrophoresis were then pooled into 47 mixed samples (eight or nine amplicons in each sample) with different barcode combinations as the short-gun sequencing library. The paired-end sequencing was performed on the MiSeq PE250 platform (Illumina, San Diego, US). Pairs of reads from the original DNA fragments were merged using FLASH [41], and filtered by QIIME [42]. The PCR amplicons from two 96-plates were pooled in one sequencing run to obtain a coverage of 50,000–100,000 reads per amplicon.

2.4. MHC Allele Genotyping and Classification

After amplicon sequencing, Trimmomatic v 0.36 [43] was used to filter the sequencing data from genomic DNA and RNA samples. In this step, (1) adapter sequences; (2) the reads that more than 10% sites were missing and (3) the reads more than 50% sites were lower than Q20 were removed. All qualified reads were uploaded to the Amplisas server [44,45]

for MHC allele genotyping. Amplisas is an online server for analysing amplicon sequencing result that integrates demultiplexing, variant clustering and allele filtering by user-specified parameters. Reads were first assigned to individuals by the unique barcode combinations and then clustered by default parameters. Amplisas includes a set of rules that aid in the detection and removal of PCR artefacts, which are prone to occur during multi-locus amplification and are a common problem in any MHC genotyping approach [46]. To further separate sequencing artefacts from putatively true alleles, we manually deleted some alleles based on the following two criteria: (1) remove variants whose maximum per amplicon frequency depth (MPAF) was lower than 1%; (2) exclude variants that appeared only in a single individual in the dataset. Given that MHC genes often exhibit an excess of rare alleles, the latter criterion might bias the retained allele repertoire towards higher-frequency alleles. However, we here chose this approach to achieve higher confidence in the allele calls and also because of the large sample size. For other studies, corresponding filtering settings need to be chosen carefully with regard to the sample size and the specific analyses that are intended.

2.5. Functional Allele Identification and Phylogenetic Analysis

Previous studies on MHC have revealed three broad classes of allelic sequences that can be detected by amplicon sequencing: (1) non-functional alleles, which are characterized by premature stop-codons or other features that prevent their translation and which are usually derived from MHC pseudogenes that are often remnant of the birth-and-death model by which the MHC gene family is thought to evolve [47]. (2) non-classical alleles, which are characterized by low polymorphism and no or low expression as well as a lack of signatures of selection, presumed to derive from genes coding for non-classical MHC molecules that serve basal functions in antigen-loading, and (3) classical alleles, which are characterized by high sequence polymorphism, signatures of selection and moderate to high expression. We therefore first examined whether alleles bear complete open reading frame by sequence alignment and those who do not bear a complete open reading frame were considered as non-functional alleles. Previous studies in passerines have found classical as well as substantial numbers of non-classical MHC alleles [22,45]. Compared with classical alleles, non-classical ones have lower expression levels and different evolutionary histories. We therefore carried out an additional filter step based on phylogenetic analysis and expression level to investigate the possibility of non-classical alleles in our study system. The relative expression level ($D_{exp-depth}$) of each MHC I allele was measured as relative read depth per allele ($D_{exp-depth}$ = the sequencing depth of the allele ÷ the sequencing depth of the corresponding individual). This analysis identified a highly supported monophyletic clade that only included non-expressed alleles and was thus designated as non-classical clade and excluded from further analysis.

To infer the phylogeny of MHC alleles, nucleotide substitution models were first tested with jModelTest (Version 2.1.3) and the best-fit model (TN93+I+G) was determined based on Bayesian Information Criterion (BIC). The Bayesian tree was then reconstructed using BEAST (Version1.8.0) and the Markov chain Monte Carlo (MCMC) was set with the length of the chain as 1×10^7 and log parameters as every 1×10^3 generations. Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/, accessed on 17 April 2021) was used to measure the convergence of MCMC chains until the ESS of all parameters were above 200. Finally, the maximum credibility tree was estimated by TreeAnnotator (Version 1.8.0) and the selected tree was adjusted and analyzed with iTOL (Version 6.6). Two phylogenetic trees were constructed for all 478 alleles and alleles with 212 bp and 215 bp length.

2.6. Historical Selection on MHC Alleles

To examine positively selected sites (PSSs) in classical MHC genes, we calculated the ratio of nonsynonymous (dN) to synonymous sites (dS) by CodeML, implemented in PAML [48]. A dN/dS ratio > 1 indicates positive selection whereas ratios < 1 indicates negative (purifying) selection. Here, we used site models implemented in CodeML that

allow dN/dS to vary across amino acid sites to examine which sites are under positive selection. Two hypotheses of codon evolution were tested here, using two pairs of models in CodeML. Models M1a and M7 represent nearly neutral codon substitution models, while M2a and M8 are positive selection codon substitution models that allow sites to have dN/dS > 1. A likelihood ratio test was used to identify the model that better supports the data (M1a vs. M2a and M7 vs. M8). The Bayes empirical Bayes (BEB) was implied to identify PSSs that have a dN/dS ratio significantly above 1 [49]. Finally, we calculated the nucleotide diversity of PSS, non-PSS, and all sites together.

2.7. Supertype Identification

The amino acid residues coding for the peptide-binding region (PBR) of the MHC molecule are expected to be most variable across MHC variants and usually bear signatures of strong positive selection, as they are largely responsible for the specific binding and presenting of pathogen-derived peptides. However, MHC research in non-model species suffers from the difficulty to accurately identify PBR sites, so PSSs are commonly used as a proxy for characterizing the specific binding properties of MHC alleles instead. In humans, the PBR and PSS sets do indeed overlap substantially [50,51]. MHC supertypes are defined by clustering MHC variants with similar binding properties based on the physicochemical properties of individual amino acid sites that are in contact with antigens [52]. In order to identify supertypes, groups of MHC variants that share similar binding properties, we first characterized the physicochemical properties of PSS amino acid by five descriptor values hydrophobicity (z1), steric bulk (z2), polarity (z3), and electronic effects (z4 and z5) [52,53] and then translated into a matrix. This matrix was then used for K-mean clustering algorithm and discriminant analysis of principal components (DAPC) using "adegenet" package in R [54]. The optimal number of supertypes was determined by Bayesian Information Criterion score. Once the number of clusters was chosen, DAPC was applied to visualize the relationship between supertypes by drawing a scatterplot using the first two PCs [55].

2.8. Avian Malaria Identification

Avian malaria parasites were identified through nested PCR amplification of a partial cyt b gene from the parasite's mitochondrial genome in 130 Godlewski's buntings. (35 in 2015 and 95 in 2016) [37]. The full lab protocols and bioinformatics methods were described in the previous study [37]. The pattern of malaria infection in terms of dominant lineages was consistent with the previous study. Therefore, combined with data generated in the previous study, a total of 306 buntings were screened for malaria parasites between 2014 and 2016 (35 in 2014, 176 in 2015, 95 in 2016).

2.9. Association between MHC Variation and Malaria Parasites

We examined the association between MHC class I variation and four dominant malaria lineages (DENVID02, ALARV04, PADOM02 and EMSPO05) in 306 Godlewski's buntings that were both genotyped at MHC class I gene and screened for malaria parasites using generalized and general linear mixed models. First, we tested whether MHC variation is associated with the infection of each malaria lineage (0/1: not infected/infected) using generalized linear mixed models with binomial error distribution. Then, we tested whether MHC variation is associated with the infected malaria lineages using a general linear mixed model. For each model, the sampling year was included as a random effect, the number of supertypes and the absence (0) or presence (1) of each MHC supertype were included as fixed effects. In addition, we constructed another set of models to replace the number of supertypes with its quadratic term to test optimal heterozygote advantage hypothesis. All models were run in R using the package lme4 [56].

3. Results

3.1. MHC Allele Diversity and Classification

Following the Amplisas genotyping pipeline, a total of 478 unique sequence variants of MHC I Exon 3 were called from 326 buntings, 282 of that were identified in at least two buntings. There was some length variability, with mainly three lengths of the variants found in this study, 212 bp, 214 bp and 215 bp. After alignment with MHC class I alleles of great tits (*Parus major*), we found only 212 bp-long and 215 bp-long variants to bear a complete open reading frame (ORF). Therefore, the other alleles without a complete ORF were considered as non-functional alleles and removed from further analysis. In summary, a total of 184 alleles satisfied the initial filter requirements for functional alleles. An average of 86.7% of the alleles was shared between the technical replicates (n = 22) included in sequencing and genotyping. The maximum number of alleles per individual is 30 with a mean of 16.1 ± 3.3 alleles per individual, suggesting copy number variation of the MHC class I genes, with up to at least 15 loci in Godlewski's bunting. There were several monophyletic clades identified in the Bayesian tree (Figure 1) for the 184 functional alleles with a length of 212 bp and 215 bp. We found a total of twelve expressed MHC I alleles with the $D_{exp-depth}$ varying from 0.04 to 0.58. These expressed alleles distribute throughout the tree, except in Clade 1 (Figure 1, Supplementary Materials Table S1). This monophyletic clade (Clade 1) is composed of only 215 bp alleles and exhibits no alleles with evidence of expression (Figure 1). In addition, we also found that clade 1 clustered with the 214 bp alleles that were identified as non-functional with a high posterior probability (p = 0.94) in a phylogenetic tree using all 478 primarily identified sequence variants (Supplementary Materials Figure S2). Therefore, alleles in clade 1 were considered as non-classical alleles (Supplementary Materials Figure S3), leaving a total of 160 classical alleles identified in the investigated buntings (Supplementary Materials Figure S4). The number of classical alleles varied among individuals from one to eighteen with an average number of 8.4 ± 2.3 classical alleles per individual. The maximum number of classical alleles per individual suggested that there are up to nine classical MHC class I loci. From these classical alleles, nine supertypes (Figure 2) were identified in the population with a mean of 5.0 ± 1.3 supertypes per individual (Table 1).

Table 1. MHC diversity of MHC Class I exon 3 in Godlewsiki's Bunting.

	Range	Mean	Median	SD	Minimum Number of Loci
Number of alleles per individual	1–30	16.06	16	3.33	15
Number of classical alleles per individual	1–18	8.42	8	2.33	9
Number of supertypes per individual	1–8	4.96	5	1.26	NA

3.2. Historical Selection Analysis

Between the nearly neutral models (M1/M7) and positive selection models (M2/M8), the positive selection models fitted the data better in PAML (Table 2). Ten positively selected sites (Table 3) were identified by Bayes empirical analysis in both models (M2 and M8). All of these ten sites were segregating and eight of them were highly variable, bearing at least four different amino acids. Compared with all sites ($\pi = 0.147$) and only non-positively selected sites ($\pi = 0.106$), positively selected sites have an outstanding high nucleotide diversity ($\pi = 0.481$) (Figure 3).

3.3. Association between MHC Variation and Malaria Parasites

We found that specific MHC supertypes were nominally associated with two lineages of malaria parasite, ALARV04 and EMSPO05. Supertype 3 and 8 was nominally associated with increased and decreased infection probability of ALARV04 respectively. Supertype 1 was nominally associated with increased infection probability of EMSPO05 (Table 4). However, the associations would not be significant if multiple-testing correction would be applied. Neither DENVID02 nor PADOM02 were associated with specific MHC supertypes

(Supplementary Materials Table S2). Finally, MHC diversity measured as the number of MHC supertypes was not associated with any malaria lineages.

Specific MHC supertypes were associated with ALARV04 and EMSPO05 separately. We did not find association between the quadratic term of MHC diversity and malaria lineage (Supplementary Materials Table S3).

Table 2. Results of PAML analyses testing for selection on classical MHC genes.

Model	Parameter Estimates	Model Comparison	*p*(Δ*LTR*)
M0	ω0 = 0.54248		
M1	ω0 = 0.08505 p0 = 0.63235 ω1=1.000 p1 = 0.36765		
M2	ω0 = 0.04618 p0 = 0.47820 ω1 =1.000 p1 = 0.36387 ω2 =3.37284 p2 = 0.15794	M1 vs. M2	<0.05
M7	p = 0.19197 q = 0.29164		
M8	p0 = 0.83551 p = 0.24542 q = 0.54364 (p1 = 0.16449) ω = 2.78333	M7 vs. M8	<0.05

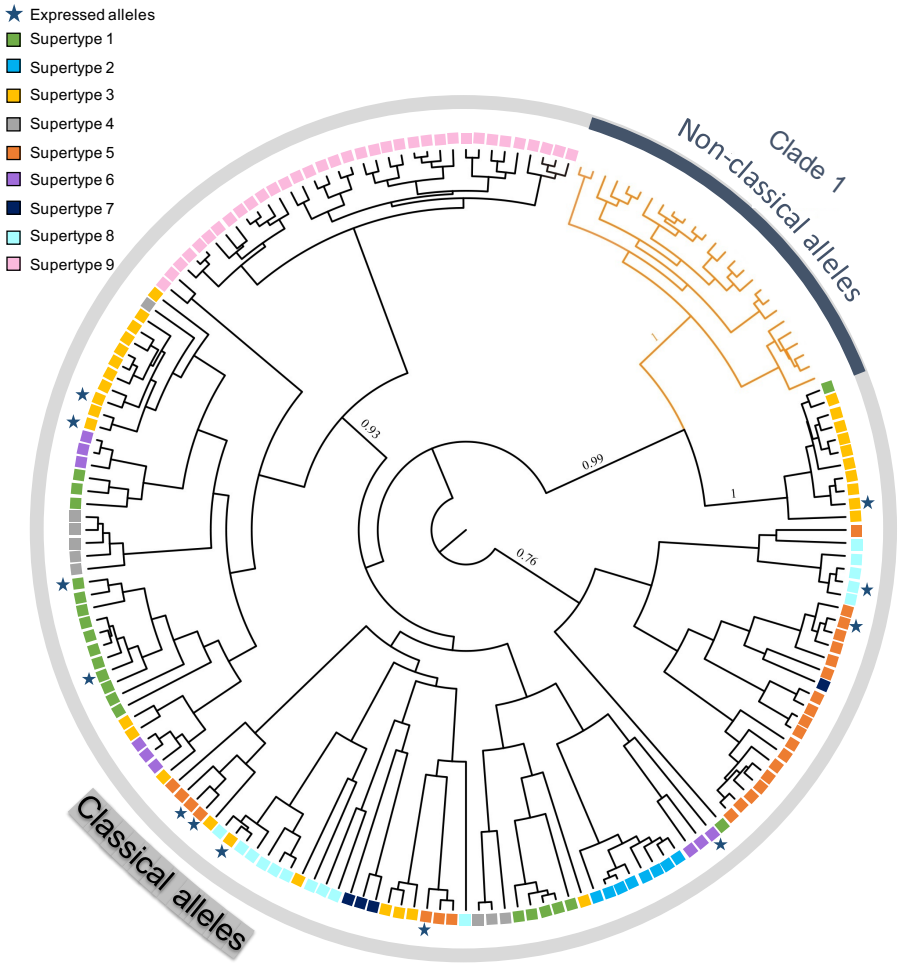

Figure 1. Bayesian phylogenetic tree depicting the major subset of MHC alleles in Godlewski's bunting. Asterisks indicate the expressed alleles. Posterior probabilities higher than 0.70 are shown by the node. The clade 1 in orange with high support without any expressed alleles was identified as non-classical clade.

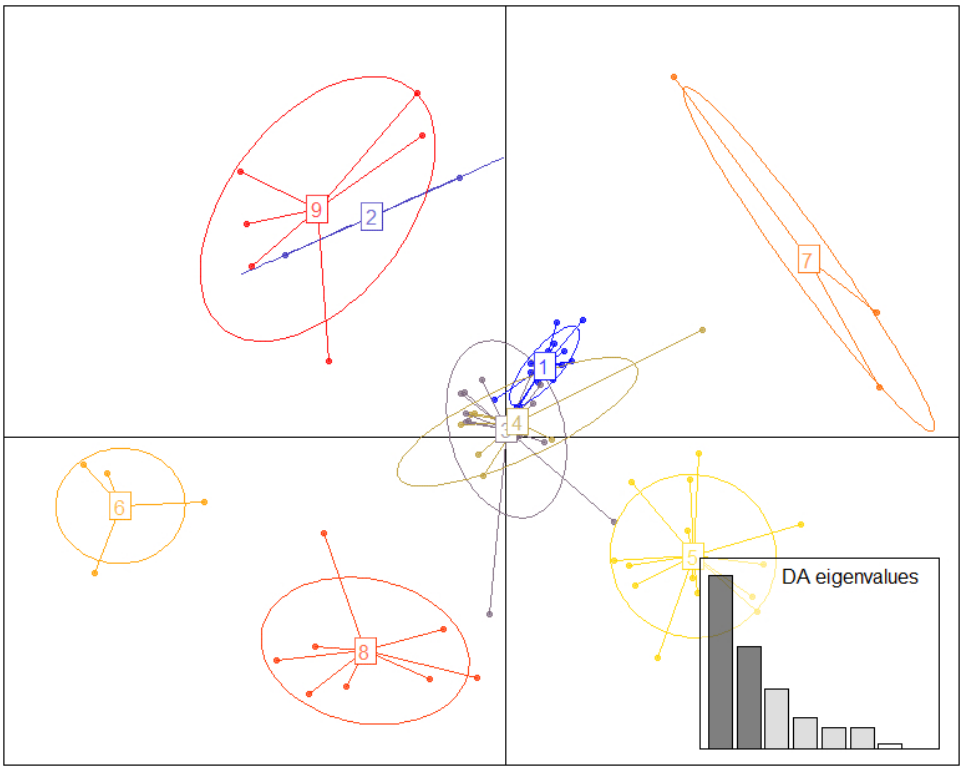

Figure 2. DAPC scatterplot of the nine MHC class I supertypes. Seven Principal components (PCs) were retained during analyses to describe the relationship between the clusters. The scatterplot shows only the first two PCs of the DAPC of MHC supertypes. The bottom right graph illustrates the variation explained by the PCs. Alleles are represented as dots and supertypes as ellipses.

Figure 3. Positive selection on amino acid sites. A sequence logo showing the relative frequencies of amino acids of the fragment of the MHC class I exon 3. The plot is based on sequences of all putative expressed alleles. Sites under positive selection are indicated with asterisks for sites with high polymorphism and dots for sites with low polymorphism.

Table 3. Ten positively selected sites detected by the Bayes empirical Bayes (*: $p > 95$ %; **: $p > 99$%).

Site	Model 2		Model 8	
	Pr ($\omega > 1$)	**Post Mean ± SE**	**Pr ($\omega > 1$)**	**Post Mean ± SE**
5	0.990 *	3.479 ± 0.266	0.999 **	2.502 ± 0.069
6	0.980 *	3.454 ± 0.363	0.998 **	2.499 ± 0.097
9	1.000 **	3.505 ± 0.095	1.000 **	2.503 ± 0.055
11	0.951 *	3.380 ± 0.544	0.998 **	2.499 ± 0.095
34	0.799	2.999 ± 1.004	0.978 *	2.464 ± 0.258
52	1.000 **	3.505 ± 0.095	1.000 **	2.503 ± 0.055
53	0.760	2.899 ± 1.068	1.000 **	2.502 ± 0.064
60	1.000 **	3.505 ± 0.095	1.000 **	2.503 ± 0.055
66	0.999 **	3.503 ± 0.112	1.000 **	2.503 ± 0.056
67	1.000 **	3.505 ± 0.096	1.000 **	2.503 ± 0.055

Table 4. The association between MHC variation and malaria lineages. Nominally significant results ($p < 0.05$) are marked in Bold.

Fixed Effects	Estimate	Standard Error	Z Value	*p* Value	Malaria Lineage
Intercept	0.488	1.610	0.303	0.762	ALARV04
Supertype 1	-0.367	0.262	−1.403	0.161	
Supertype 2	-0.153	0.279	−0.550	0.583	
Supertype 3	**1.211**	**0.593**	**2.044**	**0.041**	
Supertype 4	0.067	0.251	0.266	0.790	
Supertype 5	0.034	0.260	0.132	0.895	
Supertype 6	−0.319	0.245	−1.304	0.192	
Supertype 7	0.074	0.542	0.137	0.891	
Supertype 8	**−0.613**	**0.307**	**−1.994**	**0.046**	
Supertype 9	−0.119	1.574	−0.076	0.940	
Number of supertypes	−0.161	0.100	−1.612	0.107	
Intercept	−1.912	1.712	−1.117	0.264	EMSPO05
Supertype 1	**1.628**	**0.622**	**2.618**	**0.009**	
Supertype 2	−0.002	0.412	−0.006	0.995	
Supertype 3	−0.397	0.633	−0.627	0.531	
Supertype 4	0.250	0.374	0.668	0.504	
Supertype 5	0.690	0.436	1.582	0.114	
Supertype 6	0.475	0.367	1.295	0.195	
Supertype 7	0.298	0.814	0.367	0.714	
Supertype 8	−0.136	0.446	−0.305	0.761	
Supertype 9	−2.341	1.580	−1.482	0.138	
Number of supertypes	0.107	0.148	0.723	0.470	

4. Discussion

4.1. MHC Variation in Godlewski's Bunting

Passerine birds have been reported to have complex MHC systems with a large number of MHC alleles (and thus gene copies) as well as extensive existence of pseudogenes [17]. Prior to the usage of next generation sequencing (NGS) in MHC genotyping, the identification of passerine MHC genes was hampered significantly by high cost and incomplete allele identification because of limited sequencing depth. With the advent of NGS technology, an increasing number of studies have started applying this methodology to genotype passerine MHC genes, yielding improved results with lower cost and a higher number of alleles identified [21,25–29]. It is still a matter of discussion whether the increase in number of alleles per individual, which is often observed upon switching from older methods to NGS-based protocols, indicates that some alleles were previously missed, or whether NGS-based methods are more prone to overestimating allele number [26,57]. However, NGS technology is undoubtedly facilitating MHC genotyping in many species that could otherwise not be studied in this context, and this is particularly true for passerine birds. In this study, we established an NGS-based protocol to successfully genotype the MHC class I genes in Godlewski's bunting and identified 160 functional alleles in a large population of the buntings. The overall repeatability of the allele calling in our dataset was 86.7%. Similar levels of repeatability have also been reported in other NGS-based genotyping studies and the inconsistency of results may either come from "allelic dropout", i.e., missed alleles due to the stochastic nature of PCR amplification [45], or from mis-incorporation of PCR artefacts, which are prone to occur specifically with high PCR cycle numbers, such as used here. Future genotyping efforts should thus aim to minimize the number of PCR cycles further and also consider employing other measures to avoid artefact formation [46].

The average number of functional alleles per individual (8.4 ± 2.3) in Godlewski's buntings falls into the range of individual MHC variability in other passerine birds, such as rufous-collared sparrow (*Zonotrichia capensis*) (4.4 ± 1.8) [29], common yellowthroat

(*Geothlypis trichas*) (8.4 ± 0.4) [28], house sparrow (*Passer domesticus*) (4.7 ± 1.5) [58] and scarlet rosefinch (*Carpodacus erythrinus*) (8.0 ± 1.6) [26] which were also genotyped by NGS-based methods, but see great tit (23.8 ± 3.9) [27]. Recently, the number of MHC loci was predicted in large number of passerine species [18]. Compared with the number of MHC loci predicted in other passerine birds, Godlewski's bunting has a high level of individual diversity at MHC class I genes.

4.2. Classical and Non-Classical MHC Alleles

Among the initially identified MHC class I sequences, we found a substantial number of non-functional sequence variants. These variants were readily identified by lacking an open reading frame, and are likely representing MHC pseudogenes, i.e., remnants of the birth-and-death process of MHC gene evolution [47]. Such non-functional sequence variants have also been detected in previous studies of other passerines [27,29]. More interestingly, we also found a clear pattern of subdivision among the putatively functional MHC class I alleles in Godlewski's bunting. A group of alleles clustered in a monophyletic clade with shallow branches. Analysis of MHC class I gene expression in seven individuals corroborated that those putatively non-classical alleles exhibit no expression in blood. Indeed, a similar pattern has been reported in a previous study, revealing the existence of both classical and non-classical MHC I alleles in house sparrows. Compared to classical MHC alleles, non-classical alleles appear to have a different evolutionary history and exhibit limited expression [22]. Thus, our results largely reflected the existence of classical and non-classical MHC alleles in Godlewski's buntings.

The expressed alleles were detected in seven adult male buntings from the western mountain area of Beijing. These alleles were widely distributed in different highly supported clades except clade 1 in the phylogenetic tree (Figure 1). Due to the limited availability of RNA samples, the expression of MHC I alleles is rarely studied comprehensively in passerine birds. However, without using expression data and careful phylogenetic analysis, these patterns might be missed. This could in turn lead to inaccurate identification and classification of MHC alleles and thus overestimate the functional MHC diversity and bias downstream analysis. Here, our results emphasize the importance of using integrative methods, which combined phylogenetic reconstruction and gene expression analysis to identify classical MHC alleles. Further effort should be made to collect female and juvenile samples for a more comprehensive understanding of the MHC diversity.

4.3. Selection on MHC Genes in Godlewski's Buntings

In this study, ten amino acid residues in the exon 3 of MHC class I genes of Godlewski's buntings showed signatures of historical positive selection. Among these positive selected sites, eight of them exhibited a particularly high number of variants (more than four different amino acids). In addition, nucleotide diversity was found to be comparatively higher in positively selected sites. These patterns indicate that MHC class I genes in Godlewski's buntings evolve under strong balancing selection which are consistent with studies of other passerine birds [27,29,39].

Apart from signatures of historical selection, we also found potential evidence of contemporary selection. Specific supertypes were nominally associated with two dominant lineages while the linear and quadratic term of MHC diversity, measured as total number of supertypes were not associated with any dominant malaria lineage. Such results suggest negative frequency-dependent selection or fluctuating selection rather than heterozygote advantage might be acting. Among these malaria-associated MHC supertypes, one of them was negatively associated with the infection probability of malaria lineages while others were positively associated with the infection probability. Previous studies have demonstrated the positive association between MHC alleles or supertypes can result from quantitative resistance to malaria infection [33,59]. However, the intensity of malaria infection in Godlewski's buntings was not studied. We cannot determine whether the supertypes that were positively associated with infection probability of malaria lineages were

advantageous supertypes conferring quantitative resistance or real susceptible supertypes. Future work on our study system could thus focus on testing associations between intensity of malaria infection and MHC alleles or supertypes to determine whether specific alleles confer susceptibility or quantitative resistance for the host in this species.

5. Conclusions

In conclusion, we genotyped the MHC class I variation of large number of Godlewski's buntings using amplicon sequencing. We demonstrated that Godlewski's buntings bear high level of MHC diversity across the passerines. In addition, we also identified signatures of both historical and potential contemporary selection on MHC genes. The selection mechanism could be negative frequency-dependent selection or fluctuating selection. Our results suggest that examining associations between MHC variation and malaria parasites in passerine birds have great potential for studying selection on MHC genes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d14110925/s1, Figure S1: Map of the study region and sample site distribution of Godlewski's Bunting; Figure S2: Phylogenetic tree constructed using 478 alleles; Figure S3: Amino acid variations of putatively non-classical MHC I alleles which clustered in clade 1; Figure S4: Amino acid variations in identified classical MHC I alleles in Godlewski's bunting; Table S1: The Genbank accession numbers of twelve expressed alleles; Table S2: The association between MHC supertypes and infection of malaria lineages when linear term of MHC diversity fitted; Table S3: The association between MHC supertypes and infection of malaria lineages when linear and quadratic term of MHC diversity fitted.

Author Contributions: Formal analysis, W.H.; investigation, W.H., X.W., B.L., Y.P. and L.D.; data curation, W.H. and X.W.; Visualization, W.H. and X.W.; writing-original draft, W.H.; writing—review & editing, X.W., T.L.L. and L.D.; funding acquisition, L.D. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 31772444 and No. 32170418 to LD) and the Monitoring Fund for the Haemosporidian Parasites by the National Forestry and Grassland Administration of China.

Institutional Review Board Statement: All animal experiments were carried out in accordance with the guidelines and approval of the Animal Research and Ethics Committee of College of Life Science,Beijing Normal University (permissions, CLS-EAW-2013-007).

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data are available from the Genbank with accession OP173808-OP174285.

Acknowledgments: We thank Jiaxin Cao for the technique support of RNA extraction and Beijing Baihuashan National Natural Reserve for assistance in fieldwork.

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

References

1. Pilosof, S.; Fortuna, M.A.; Cosson, J.F.; Galan, M.; Kittipong, C.; Ribas, A.; Segal, E.; Krasnov, B.R.; Morand, S.; Bascompte, J. Host–parasite network structure is associated with community-level immunogenetic diversity. *Nat. Commun.* **2014**, *5*, 5172. [CrossRef] [PubMed]
2. Bernatchez, L.; Landry, C. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *J. Evol. Biol.* **2003**, *16*, 363–377. [CrossRef] [PubMed]
3. Wilson, A.J.; Reale, D.; Clements, M.N.; Morrissey, M.M.; Postma, E.; Walling, C.A.; Kruuk, L.E.B.; Nussey, D.H. An ecologist's guide to the animal model. *J. Anim. Ecol.* **2010**, *79*, 13–26. [CrossRef]
4. Piertney, S.B.; Oliver, M.K. The evolutionary ecology of the major histocompatibility complex. *Heredity* **2006**, *96*, 7–21. [CrossRef]
5. Edwards, S.V.; Hedrick, P.W. Evolution and ecology of MHC molecules: From genomics to sexual selection. *Trends Ecol. Evol.* **1998**, *13*, 305–311. [CrossRef]
6. Spurgin, L.G.; Richardson, D.S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. R. Soc. B-Biol. Sci.* **2010**, *277*, 979–988. [CrossRef]

7. Radwan, J.; Babik, W.; Kaufman, J.; Lenz, T.L.; Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet.* **2020**, *36*, 298–311. [CrossRef]
8. Takahata, N.; Nei, M. Allelic Genealogy under Overdominant and Frequency-Dependent Selection and Polymorphism of Major Histocompatibility Complex Loci. *Genetics* **1990**, *124*, 967–978. [CrossRef]
9. Bodmer, W.F. Evolutionary Significance of the HL-A System. *Nature* **1972**, *237*, 139–145. [CrossRef]
10. Slade, R.W.; Mccallum, H.I. Overdominant Vs Frequency-Dependent Selection at Mhc Loci. *Genetics* **1992**, *132*, 861–862. [CrossRef]
11. Hedrick, P.W. Pathogen resistance and genetic variation at MHC loci. *Evolution* **2002**, *56*, 1902–1908. [CrossRef] [PubMed]
12. Milinski, M. The major histocompatibility complex, sexual selection, and mate choice. *Annu. Rev. Ecol. Evol. Syst.* **2006**, *37*, 159–186. [CrossRef]
13. Sin, Y.W.; Annavi, G.; Newman, C.; Buesching, C.; Burke, T.; Macdonald, D.W.; Dugdale, H.L. MHC class II-assortative mate choice in European badgers (Meles meles). *Mol. Ecol.* **2015**, *24*, 3138–3150. [CrossRef] [PubMed]
14. Huchard, E.; Knapp, L.A.; Wang, J.L.; Raymond, M.; Cowlishaw, G. MHC, mate choice and heterozygote advantage in a wild social primate. *Mol. Ecol.* **2010**, *19*, 2545–2561. [CrossRef]
15. Paterson, S.; Pemberton, J.M. No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc. R. Soc. B-Biol. Sci.* **1997**, *264*, 1813–1819. [CrossRef]
16. Kaufman, J.; Milne, S.; Gobel, T.W.F.; Walker, B.A.; Jacob, J.P.; Auffray, C.; Zoorob, R.; Beck, S. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **1999**, *401*, 923–925. [CrossRef]
17. Westerdahl, H. Passerine MHC: Genetic variation and disease resistance in the wild. *J. Ornithol.* **2007**, *148*, S469–S477. [CrossRef]
18. Minias, P.; Pikus, E.; Whittingham, L.A.; Dunn, P.O. Evolution of Copy Number at the MHC Varies across the Avian Tree of Life. *Genome Biol. Evol.* **2019**, *11*, 17–28. [CrossRef]
19. He, K.; Minias, P.; Dunn, P.O. Long-Read Genome Assemblies Reveal Extraordinary Variation in the Number and Structure of MHC Loci in Birds. *Genome Biol. Evol.* **2021**, *13*, evaa270. [CrossRef]
20. O'Connor, E.A.; Strandh, M.; Hasselquist, D.; Nilsson, J.A.; Westerdahl, H. The evolution of highly variable immunity genes across a passerine bird radiation. *Mol. Ecol.* **2016**, *25*, 977–989. [CrossRef]
21. Biedrzycka, A.; O'Connor, E.; Sebastian, A.; Migalska, M.; Radwan, J.; Zajac, T.; Bielanski, W.; Solarz, W.; Cmiel, A.; Westerdahl, H. Extreme MHC class I diversity in the sedge warbler (Acrocephalus schoenobaenus); selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evol. Biol.* **2017**, *17*, 1–12. [CrossRef] [PubMed]
22. Drews, A.; Strandh, M.; Raberg, L.; Westerdahl, H. Expression and phylogenetic analyses reveal paralogous lineages of putatively classical and non-classical MHC-I genes in three sparrow species (Passer). *BMC Evol. Biol.* **2017**, *17*, 152. [CrossRef] [PubMed]
23. Babik, W.; Taberlet, P.; Ejsmond, M.J.; Radwan, J. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol. Ecol. Resour.* **2009**, *9*, 713–719. [CrossRef] [PubMed]
24. Babik, W. Methods for MHC genotyping in non-model vertebrates. *Mol. Ecol. Resour.* **2010**, *10*, 237–251. [CrossRef] [PubMed]
25. Zagalska-Neubauer, M.; Babik, W.; Stuglik, M.; Gustafsson, L.; Cichon, M.; Radwan, J. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evol. Biol.* **2010**, *10*, 1–15. [CrossRef] [PubMed]
26. Promerova, M.; Babik, W.; Bryja, J.; Albrecht, T.; Stuglik, M.; Radwan, J. Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine-CE-SSCP and 454 pyrosequencing. *Mol. Ecol. Resour.* **2012**, *12*, 285–292. [CrossRef]
27. Sepil, I.; Moghadam, H.K.; Huchard, E.; Sheldon, B.C. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol. Biol.* **2012**, *12*, 1–19. [CrossRef]
28. Dunn, P.O.; Bollmer, J.L.; Freeman-Gallant, C.R.; Whittingham, L.A. Mhc Variation Is Related to a Sexually Selected Ornament, Survival, and Parasite Resistance in Common Yellowthroats. *Evolution* **2013**, *67*, 679–687. [CrossRef]
29. Jones, M.R.; Cheviron, Z.A.; Carling, M.D. Variation in positively selected major histocompatibility complex class I loci in rufous-collared sparrows (Zonotrichia capensis). *Immunogenetics* **2014**, *66*, 693–704. [CrossRef]
30. Bensch, S.; Hellgren, O.; Perez-Tris, J. MalAvi: A public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Mol. Ecol. Resour.* **2009**, *9*, 1353–1358. [CrossRef]
31. Clark, N.J.; Clegg, S.M.; Lima, M.R. A review of global diversity in avian haemosporidians (Plasmodium and Haemoproteus: Haemosporida): New insights from molecular data. *Int. J. Parasitol.* **2014**, *44*, 329–338. [CrossRef] [PubMed]
32. Biedrzycka, A.; Bielanski, W.; Cmiel, A.; Solarz, W.; Zajac, T.; Migalska, M.; Sebastian, A.; Westerdahl, H.; Radwan, J. Blood parasites shape extreme major histocompatibility complex diversity in a migratory passerine. *Mol. Ecol.* **2018**, *27*, 2594–2603. [CrossRef] [PubMed]
33. Sepil, I.; Lachish, S.; Hinks, A.E.; Sheldon, B.C. Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proc. R. Soc. B-Biol. Sci.* **2013**, *280*, 20130134. [CrossRef] [PubMed]
34. Westerdahl, H.; Waldenstrom, J.; Hansson, B.; Hasselquist, D.; von Schantz, T.; Bensch, S. Associations between malaria and MHC genes in a migratory songbird. *Proc. Biol. Sci.* **2005**, *272*, 1511–1518. [CrossRef] [PubMed]
35. Jones, M.R.; Cheviron, Z.A.; Carling, M.D. Spatially variable coevolution between a haemosporidian parasite and the MHC of a widely distributed passerine. *Ecol. Evol.* **2015**, *5*, 1045–1060. [CrossRef]
36. Loiseau, C.; Zoorob, R.; Robert, A.; Chastel, O.; Julliard, R.; Sorci, G. Plasmodium relictum infection and MHC diversity in the house sparrow (Passer domesticus). *Proc. Biol. Sci.* **2011**, *278*, 1264–1272. [CrossRef]

37. Liu, B.Y.; Deng, Z.Q.; Huang, W.; Dong, L.; Zhang, Y.Y. High prevalence and narrow host range of haemosporidian parasites in Godlewski's bunting (Emberiza godlewskii) in northern China. *Parasitol. Int.* **2019**, *69*, 121–125. [CrossRef]
38. Griggio, M.; Biard, C.; Penn, D.J.; Hoi, H. Female house sparrows "count on" male genes: Experimental evidence for MHC-dependent mate preference in birds. *BMC Evol. Biol.* **2011**, *11*, 1–7. [CrossRef]
39. Alcaide, M.; Liu, M.; Edwards, S.V. Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *Peerj* **2013**, *1*, e86. [CrossRef]
40. Westerdahl, H.; Wittzell, H.; von Schantz, T.; Bensch, S. MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity* **2004**, *92*, 534–542. [CrossRef]
41. Magoc, T.; Salzberg, S.L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [CrossRef] [PubMed]
42. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [CrossRef] [PubMed]
43. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
44. Sebastian, A.; Herdegen, M.; Migalska, M.; Radwan, J. amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol. Ecol. Resour.* **2016**, *16*, 498–510. [CrossRef]
45. Biedrzycka, A.; Sebastian, A.; Migalska, M.; Westerdahl, H.; Radwan, J. Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Mol. Ecol. Resour.* **2017**, *17*, 642–655. [CrossRef]
46. Lenz, T.L.; Becker, S. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene* **2008**, *427*, 117–123. [CrossRef]
47. Nei, M.; Gu, X.; Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7799–7806. [CrossRef]
48. Yang, Z.H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
49. Yang, Z.; Wong, W.S.; Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [CrossRef]
50. Reche, P.A.; Reinherz, E.L. Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *J. Mol. Biol.* **2003**, *331*, 623–641. [CrossRef]
51. Furlong, R.F.; Yang, Z. Diversifying and purifying selection in the peptide binding region of DRB in mammals. *J. Mol. Evol.* **2008**, *66*, 384–394. [CrossRef] [PubMed]
52. Doytchinova, I.A.; Flower, D.R. In silico identification of supertypes for class II MHCs. *J. Immunol.* **2005**, *174*, 7085–7095. [CrossRef] [PubMed]
53. Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjostrom, M.; Wold, S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.* **1998**, *41*, 2481–2491. [CrossRef] [PubMed]
54. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef]
55. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 1–15. [CrossRef]
56. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2013.
57. Lighten, J.; van Oosterhout, C.; Bentzen, P. Critical review of NGS analyses for de novo genotyping multigene families. *Mol. Ecol.* **2014**, *23*, 3957–3972. [CrossRef]
58. Karlsson, M.; Westerdahl, H. Characteristics of MHC Class I Genes in House Sparrows Passer domesticus as Revealed by Long cDNA Transcripts and Amplicon Sequencing. *J. Mol. Evol.* **2013**, *77*, 8–21. [CrossRef]
59. Westerdahl, H.; Asghar, M.; Hasselquist, D.; Bensch, S. Quantitative disease resistance: To better understand parasite-mediated selection on major histocompatibility complex. *Proc. R. Soc. B-Biol. Sci.* **2012**, *279*, 577–584. [CrossRef]