

RESEARCH ARTICLE

Open Access



Balancing selection and recombination as evolutionary forces caused population genetic variations in golden pheasant MHC class I genes

Qian-Qian Zeng^{1†}, Ke He^{1,2†}, Dan-Dan Sun¹, Mei-Ying Ma¹, Yun-Fa Ge¹, Sheng-Guo Fang¹ and Qiu-Hong Wan^{1*}

Abstract

Background: The major histocompatibility complex (MHC) genes are vital partners in the acquired immune processes of vertebrates. MHC diversity may be directly associated with population resistance to infectious pathogens. Here, we screened for polymorphisms in exons 2 and 3 of the *IA1* and *IA2* genes in 12 golden pheasant populations across the Chinese mainland to characterize their genetic variation levels, to understand the effects of historical positive selection and recombination in shaping class I diversity, and to investigate the genetic structure of wild golden pheasant populations.

Results: Among 339 individual pheasants, we identified 14 *IA1* alleles in exon 2 (*IA1*-E2), 11 *IA1*-E3 alleles, 27 *IA2*-E2 alleles, and 28 *IA2*-E3 alleles. The non-synonymous substitution rate was significantly greater than the synonymous substitution rate at sequences in the *IA2* gene encoding putative peptide-binding sites but not in the *IA1* gene; we also found more positively selected sites in *IA2* than in *IA1*. Frequent recombination events resulted in at least 9 recombinant *IA2* alleles, in accordance with the intermingling pattern of the phylogenetic tree. Although some *IA* alleles are widely shared among studied populations, large variation occurs in the number of *IA* alleles across these populations. Allele frequency analysis across 2 *IA* loci showed low levels of genetic differentiation among populations on small geographic scales; however, significant genetic differentiation was observed between pheasants from the northern and southern regions of the Yangtze River. Both STRUCTURE analysis and F-statistic (F_{ST}) value comparison classified those populations into 2 major groups: the northern region of the Yangtze River (NYR) and the southern region of the Yangtze River (SYR).

Conclusions: More extensive polymorphisms in *IA2* than *IA1* indicate that *IA2* has undergone much stronger positive-selection pressure during evolution. Moreover, the recombination events detected between the genes and the intermingled phylogenetic pattern indicate that interlocus recombination accounts for much of the allelic variation in *IA2*. Analysis of the population differentiation implied that homogenous balancing selection plays an important part in maintaining an even distribution of MHC variations. The natural barrier of the Yangtze River and heterogeneous balancing selection might help shape the NYR-SYR genetic structure in golden pheasants.

Keywords: Balancing selection, *Chrysolophus pictus*, Galliformes, Genetic diversity, Major histocompatibility complex, MHC class I, Population genetics, Recombination

* Correspondence: qiuqianwan@zju.edu.cn

[†]Equal contributors

¹The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, State Conservation Center for Gene Resources of Endangered Wildlife, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

Full list of author information is available at the end of the article



Background

The major histocompatibility complex (MHC) is a primary factor in initiating immune defenses, and it is composed of glycoproteins that are specialized to present foreign antigens to T lymphocytes [1]. The MHC multigene family can be divided into at least 2 main classes of genes: class I and class II. Class I molecules are expressed in a wide variety of nucleated cells and mainly respond to intracellular parasites [2, 3]. MHC class I proteins are structurally composed of a cytoplasmic region, a transmembrane portion, and 3 extracellular domains ($\alpha 1$, $\alpha 2$, and $\alpha 3$), where antigen-binding domains ($\alpha 1$ and $\alpha 2$) encoded by exons 2–3 interact to form the hypervariable peptide-binding region (PBR) [2].

Apart from their prominent role in immune responses, much attention has been paid to MHC genes because of their extensive polymorphism. In recent years, the patterns of genetic variability in MHC class I genes have been intensively investigated across different vertebrate taxa, including mammals [4, 5], birds [6, 7], reptiles [8, 9], amphibians [10, 11], and fishes [12, 13]. The mechanisms that generate abundant MHC variation primarily involve parasite-related balancing selection [14], such as frequency-dependent selection or overdominant selection [15]. Because fast-evolving pathogens can easily escape the immune surveillance of common host MHC alleles, frequency-dependent selection prevails, contributing to the generation and retention of rare alleles [16]. Consequently, changes in the pathogen community over time and location lead to MHC variation in host populations [17]. In the case of overdominance, heterozygotes exhibiting superior recognition of a wider range of pathogens mount a stronger immune defense against pathogen infections than do homozygotes [18]. In addition, MHC-based mating preferences increase MHC gene heterogeneity in progeny [19–21].

In diverse vertebrate lineages, including birds, the MHC gene family has an extremely complex molecular architecture and genome organization with high gene copy numbers, abundant pseudogenes, frequent recombination, and homologous chromosome exchanges [22–26]. The complex evolutionary pattern in MHC genes appears to be dictated both by birth-and-death and concerted evolution [27]. In the birth-and-death mechanism, some novel duplicated genes remain functional for long periods in the genome, whereas others decay into non-functional genes (pseudogenes) or are completely eliminated from the genome [25, 28]. The concerted evolution model implies that different members of the MHC family evolve as a unit, primarily due to recombination across paralogous genes [27, 29]; as a result, genes within a species are more similar compared to orthologous genes in closely related species. Moreover, such widespread sequence homogeneity caused by repeated recombination events increases the difficulty in

assigning alleles to specific MHC loci. Therefore, given the current poor understanding of MHC genomic organization, MHC single-locus typing in some non-model avian taxa is more challenging [7, 30–32]. The chicken MHC-B (*Gallus gallus*, Phasianidae, Galliformes), the first representative of the “minimal essential” MHC, is highly condensed and dramatically simpler than the mammalian MHC, containing only 2 classical class I genes, *BF1* and *BF2* [33, 34]. The detailed knowledge of MHC organization in chickens and its compressed nature facilitated the characterization of MHC diversity at the locus-specific level [35].

Golden pheasants (*Chrysolophus pictus*, Phasianidae, Galliformes), representing an endangered species in China, are scattered in the central and western regions of Mainland China [36]. Their main distribution areas are divided by 2 major geographical barriers: the Yangtze River and Qinling-Daba Mountain (QDM) (Fig. 1). To provide efficient protective strategies for this pheasant, we prioritized screening of MHC class I variation among wild populations. Previous studies on the golden pheasant MHC have reported the complete sequence of the gold pheasant MHC-B region, which is nearly identical to the chicken MHC-B gene arrangement [37]. The 97-kb golden pheasant MHC-B comprises 20 streamlined genes and conforms to the minimal essential hypothesis. In addition, only 2 duplicated functional class I genes (*IA1* and *IA2*) were identified in this species, as in chickens [37]. Based on these findings, we examined allelic variation of classical MHC class I genes in golden pheasants. The main objectives of this study were to (i) characterize sequence polymorphisms of exons 2 and 3 from both *IA1* and *IA2* genes through locus-specific PCR amplification, (ii) examine the role of historical positive selection and recombination in shaping class I diversity, and (iii) infer the population-genetics structure on the basis of MHC variation.

Results

MHC class I diversity

For the *IA1* and *IA2* genes, no more than 2 alleles were identified per pheasant by the PCR-based, single-stranded conformation polymorphism (SSCP) technique, indicating that single-locus genotyping was optimal using the present primer sets. For *IA1*, 23 of 254 (9 %) and 54 of 236 (23 %) nucleotide sites were variable at exons 2 and 3, and 14 and 11 alleles were identified, respectively. For *IA2*, 74 of 251 (29 %) and 71 of 253 (28 %) nucleotide sites were variable, and 27 and 28 alleles were found at exons 2 and 3, respectively.

Our unpublished experimental data indicate that both *IA* genes are expressed in the golden pheasant. Through screening cDNA libraries using universal primers, 4 *IA* sequences were found for one pheasant; among them, 2

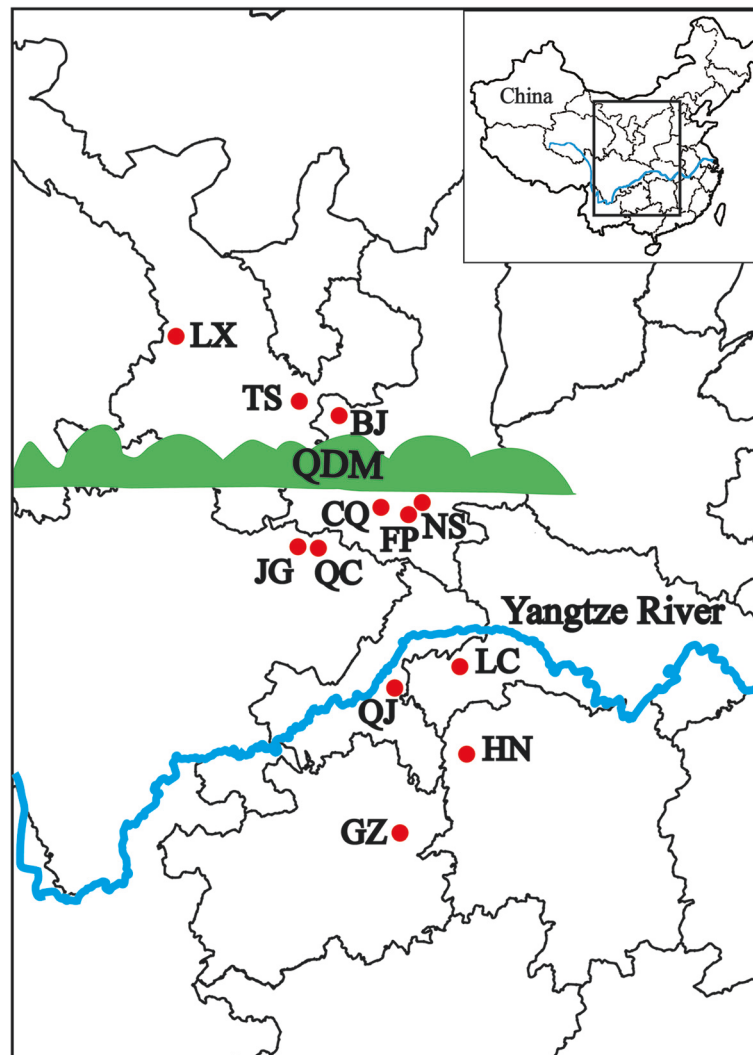


Fig. 1 Geographic sampling locations of wild golden pheasants in China. The green zone represents the Qinling-Daba Mountain (QDM), which spans across central and western China, while the blue curve indicates the Yangtze River running from east to west China. Red dots mark all wild populations sampled for our study, and the population abbreviations are as follows: LX, Linxia; TS, Tianshui; BJ, Baoji; CQ, Changqing; FP, Foping; NS, Ningshan; JG, Jiange; QC, Qingchuan; LC, Lichuan; HN: Hunan; QJ: Qianjiang; GZ, Guizhou. This figure was created based on the map of China (<https://commons.wikimedia.org/wiki/File%3AChina-map.xcf>)

sequences corresponded to alleles from *IA1* and the other 2 were from *IA2*. Moreover, none of the putative amino acid sequences showed frameshift mutations for either gene, further implying that both genes were functional. For *IA1*, 16 of 83 (19 %) and 26 of 78 (33 %) amino acid sites were variable at exons 2 and 3. Fourteen alleles at exon 2 and 11 alleles at exon 3 coded for 12 and 10 distinct amino acid sequences, respectively (Fig. 2). Remarkably, the *IA1* alleles at exon 2 (*IA1*-E2) shared a locus-specific amino acid motif constituted by 5 conserved, discontinuous residues (N74, G75, K76, S78, and D80; Fig. 2). Such a striking feature, which resembled that reported in domestic chicken *BF1* alleles [38], might be used to distinguish *IA1* from *IA2*.

For *IA2*, 38 of 83 (46 %) and 27 of 83 (33 %) amino acid sites were variable at exons 2 and 3. Twenty-seven and 28 different amino acid sequences were encoded by the exon 2 alleles (-E2) and exon 3 alleles (-E3), respectively (Fig. 2). As mentioned above, the *IA2*-E2 alleles were the most variable and diverse and showed the greatest nucleotide and amino acid distances, especially in the PBR (Table 1), whereas the *IA1*-E2 alleles were the least diverse, with both distance values in the PBR slightly less than in the non-PBR.

Selection of 2 *IA* genes

Calculations for non-synonymous (d_N : contributing to amino acid changes) and synonymous (d_S : causing no

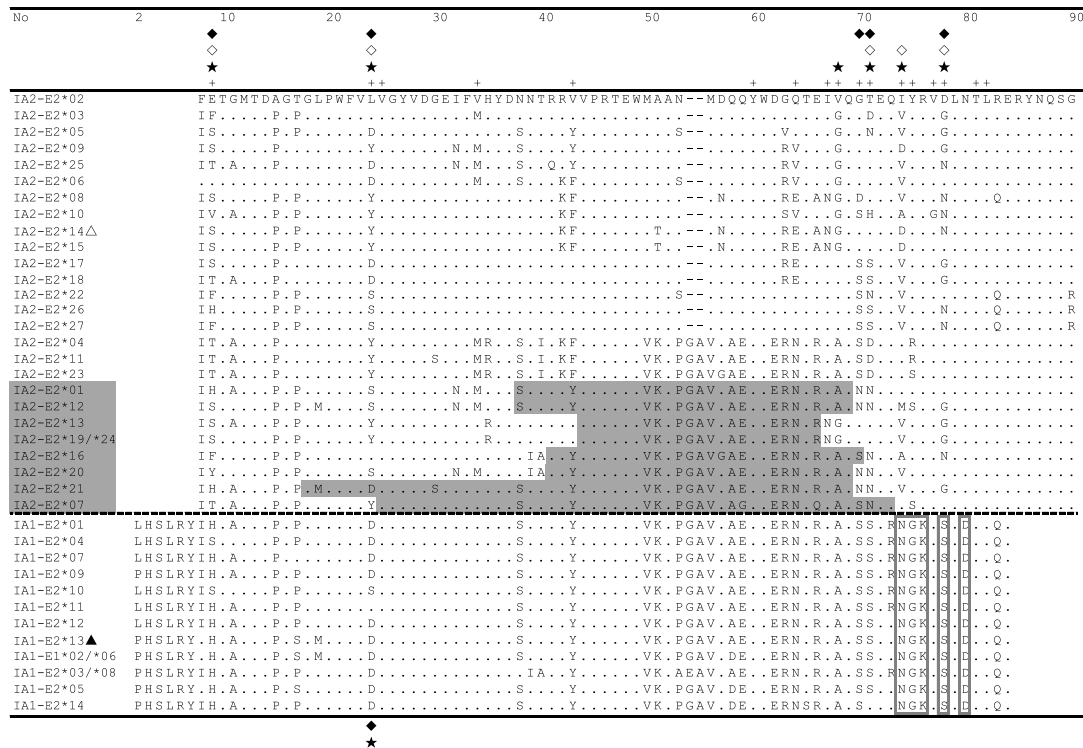
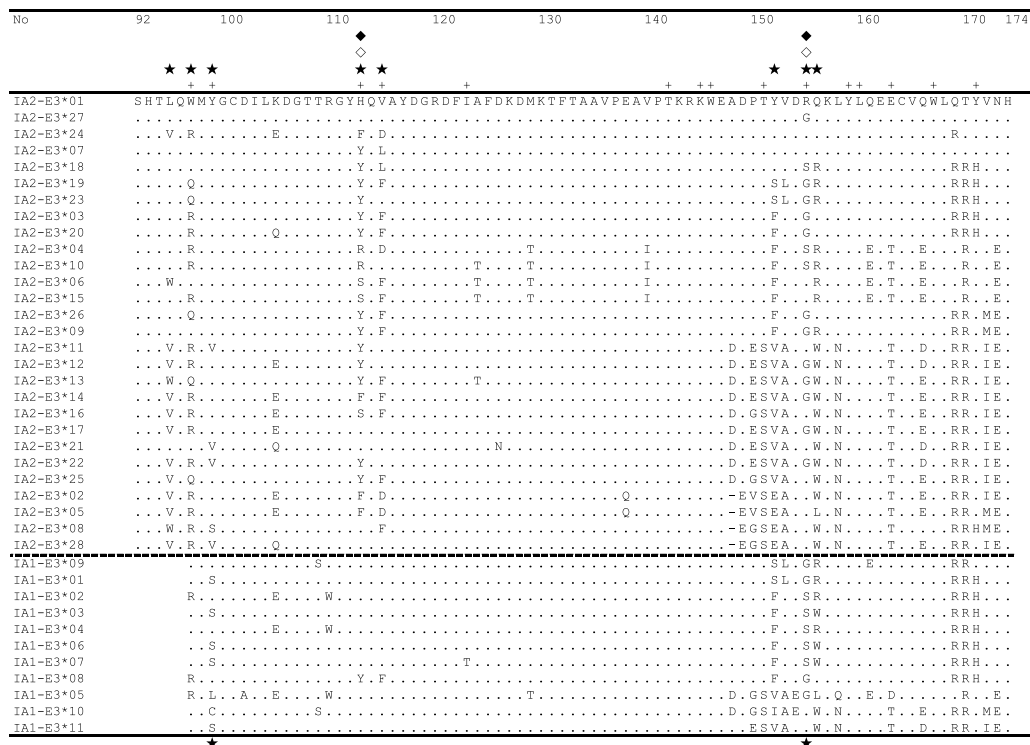
$\alpha 1$ domain $\alpha 2$ domain

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Golden pheasant *IA* (most of the $\alpha 1$ and $\alpha 2$ domains) amino acid alignment. Numbers above the alignment refer to residue positions. Locus-specific residues belonging to the *IA1* gene are indicated by open boxes. PBR sites predicted from consensus positions of chicken class I molecules [49, 50] are marked with a "+" symbol. The positively selected sites (PSSs) for *IA1* and *IA2* are indicated above and below the alignments, respectively. A filled asterisk represents PSS identified using the M8 site model, while a filled or open prism marks a PSS detected by REL or FEL, respectively. Interlocus recombination events determined by the RDP program produced 9 daughter recombinants, with their respective recombined regions indicated by grey shading; *IA2*-E2*14 as the major parent is denoted with an open triangle and *IA1*-E2*13 as the minor parent is denoted with a filled triangle

amino acid change) substitutions for *IA1* and *IA2* are shown in Table 2. Regarding the putative PBR of *IA1*-E2, d_N was almost 4-fold higher than d_S , although the difference between these 2 values was not statistically significant (Table 2). Non-PBR codons also yielded higher d_N than d_S values, but again the difference showed no significant deviation from neutrality (Table 2). Conversely, *IA2*-E2 alleles showed a significantly higher d_N than d_S in the PBR ($d_N/d_S = 9.444$; Table 2), while d_N/d_S was <1 for non-PBR codons. Surprisingly, for the $\alpha 2$ domain (E3) of both *IA* molecules, a lower d_N than d_S value was observed even in the PBR, although d_N in the PBR was approximately 2-fold higher than in the non-PBR.

Based on the alignment of 166 codons within the *IA2* gene, OmegaMap inferred 17 and 7 positively selected amino acid sites in exon 2 and exon 3, respectively. Fewer positively selected amino acids were found using other methods. Both the random-effects likelihood and fixed-effects likelihood (FEL) models identified 7 codons under positive selection (5 from exon 2, 2 from exon 3), while likelihood ratio tests, by comparison of M8 and M7 site models, detected 14 positively selected sites (6 from exon 2, 8 from exon 3; Table 3). Almost all sites under positive selection, as estimated using PAML software, were also detected using OmegaMap. Based on the combined results of the 4 methods, we concluded that 7 positive selection sites may be directly associated with peptide binding. However, there were fewer codons in the *IA1* gene under positive selection (10, 3, 1, and 0 codons found using the M8 vs. M7 in OmegaMap, PAML, REL, and FEL models, respectively; Table 3).

Phylogenetic analysis

The golden pheasant MHC class I sequences clearly clustered with published MHC I sequences from other Galliformes species, but they were separated from those of non-Galliformes birds both in the nucleic acid (Fig. 3a and c) and amino acid (Fig. 3b and d) phylogenetic trees. Within the Galliformes clade, sequences belonging to the same species tended to group together. In Fig. 3a and b, all *IA1*-E2 alleles and 12 *IA2*-E2 alleles of the golden pheasants formed an independent clade, while all *IA2*-E2 alleles fell into 2 relatively well-distinguished clusters due to extreme sequence variability. Similarly, those E3 alleles, which displayed great sequence divergences characteristic of the classical class I genes, mostly fell into 2 major clusters except for *IA1*-E3*05 and *IA1*-E3*10 (Fig. 3c and d); within one cluster, alleles from both loci comingled with each other. In addition, the network relationship among class I sequences of 7 Galliformes species also signified the absence of orthologous relationships (Additional file 1: Figure S1).

Recombination of MHC I loci in the golden pheasant

The alignment results between *IA1* and *IA2* revealed that the intermediate 44–66 amino acid residues of alleles *IA2*-E2*01, 04, 07, 11–13, 16, 19, 20, 21, 23, and 24 were nearly equivalent to the corresponding region of alleles *IA1*-E2*01, 04, 07, 09–11, and 13 (in Fig. 2), which was suggestive of intergenic recombination. This pattern was subsequently verified using at least 3 methods in the RDP software program. Those interlocus recombination events occurred intensively between *IA2*-E2*14 as the

Table 1 Nucleotide and amino acid distances of MHC I alleles for the golden pheasant

Locus	No. of alleles	K2P nucleotide distance			Poisson-corrected amino acid distance		
		All sites	PBR	Non-PBR	All sites	PBR	Non-PBR
<i>IA1</i>							
-E2 (254 bp)	14	2.9 (0.6)	2.1 (0.1)	3.1 (0.8)	6.6 (2.8)	4.4 (2.4)	7.4 (2.1)
-E3 (236 bp)	11	10.5 (1.6)	17.2 (5.4)	3.1 (0.8)	13.6 (2.8)	24.4 (8.9)	10.8 (2.9)
<i>IA2</i>							
-E2 (251 bp)	27	14.7 (1.9)	30.0 (7.5)	11.3 (1.7)	24.6 (4.1)	55.9 (21.9)	18.5 (3.7)
-E3 (253 bp)	28	14.7 (1.9)	30.0 (7.5)	11.3 (1.7)	24.6 (4.1)	55.9 (21.9)	18.5 (3.7)

Notes: Nucleotide distances were corrected for multiple substitutions with the Kimura 2-parameter model (K2P), and amino acid distances were corrected using expectations from the Poisson distribution. Distance values are given as percentages per site, with standard errors based on 1,000 bootstrap replicates (given in parentheses)

Table 2 Calculations of non-synonymous (d_N) and synonymous (d_S) substitutions for golden pheasant MHC class I genes

Locus	All sites			PBR			Non-PBR		
	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
<i>IA1</i>									
-E2	3.3 (0.8)	1.7 (0.9)	1.94	3.5 (2.4)	0.9 (0.1)	3.89	3.3 (0.9)	2.0 (1.2)	1.65
-E3	7.7 (1.5)	10.6 (2.6)	0.72	12.1 (3.9)	14.1 (7.0)	0.72	6.2 (1.7)	9.6 (2.9)	0.65
<i>IA2</i>									
-E2	13.2 (2.4)	8.5 (2.3)	1.55	34 (9.6)	3.6 (1.8)	9.44*	8.7 (1.9)	10.5 (3.2)	0.83
-E3	10.4 (2.1)	15.0 (2.8)	0.69	19.1 (6.3)	28.8 (10.7)	0.66	7.9 (2.1)	11.4 (2.9)	0.69

Notes: The d_N and d_S were estimated using the modified Nei-Gojobori algorithm. Distance values are given as percentages per site, with standard errors based on 1,000 bootstrap replicates (given in parentheses). * $P < 0.05$, as determined by performing a 1-tailed Z test

major parent (marked with Δ in Fig. 2) and *IA1*-E2*13 as the minor parent (marked with \blacktriangle in Fig. 2). They ultimately produced 9 daughter recombinants (their respective recombined regions are shown by grey shading in Fig. 2), accounting for a substantial proportion (33.3 %) of the *IA2*-E2 alleles. Thus, a genomic region (31–190 nt) of *IA2*-E2*16 was replaced respectively by its counterpart in *IA1*-E2*13 during the recombination process. Accordingly, one significant breakpoint at 109 nt ($P < 0.01$, after sequential Bonferroni correction) was detected using the GARD program. We also assessed recombination between exons 2 and 3 in *IA1* (Fig. 3e) and *IA2* (Fig. 3f), respectively. The bars (PHASE-based haplotypes) with fore-grey and after-white were the products of recombination between exon 2 (grey clusters; Fig. 3a and b) and exon 3 (white or blank clusters; Fig. 3c and d), and vice versa for the fore-white after-grey bars. Among these bars, those with a frequency higher than 0.1 should reflect inter-exon recombination hotspots.

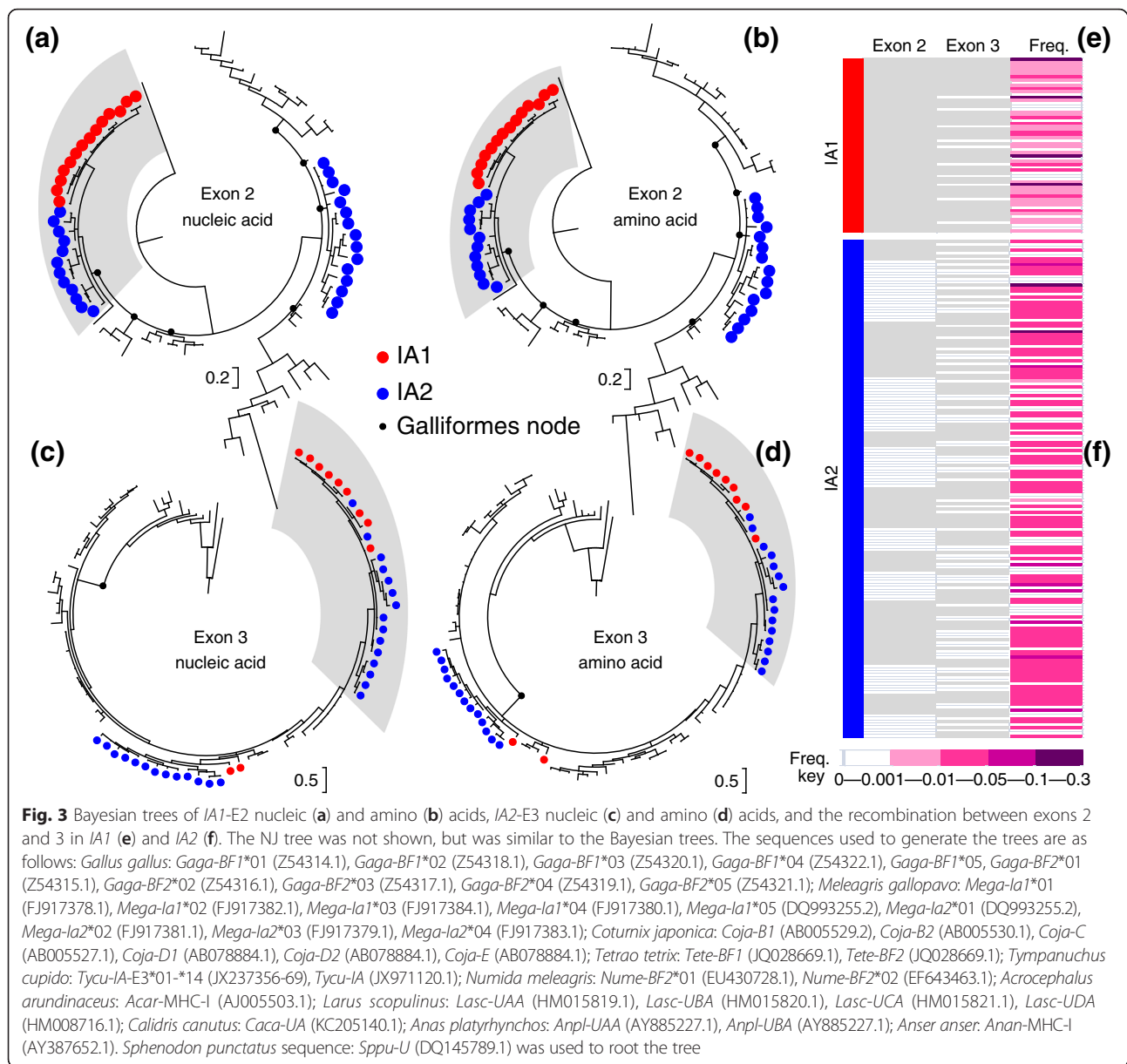
Genetic diversity and population-structure analysis of MHC I loci across golden pheasant populations

Within the populations studied, the number of *IA1* alleles ranged from 5–11 at exon 2 and from 3–9 at exon 3, whereas the number of *IA2* alleles ranged between 6 and 22 at either exon 2 or exon 3 (Additional file 2: Table S1). To handle the variation of sample sizes across the different populations, allelic richness was standardized based on a minimal sample size of 9 diploid individuals, using the rarefaction method. In this case, estimates of the allelic richness at each locus still differed among the populations (*IA1*-E2: 4.4–6.7; *IA1*-E3: 3–5.7; *IA2*-E2: 5.9–8.8; and *IA2*-E3: 5.9–9.5; Additional file 3: Table S2). Five *IA1* alleles (*IA1*-E2*03, 06 and *IA1*-E3*01, 02, 05) and 6 *IA2* alleles (*IA2*-E2*03, 04 and *IA2*-E3*01, 03, 08, 18) were observed throughout all the populations studied. In particular, 8 alleles (*IA1*-E2*03, 06, and *IA1*-E3*02, 05; *IA2*-E2*03, 04; and *IA2*-E3*01, 03) were most common, although the allele frequencies varied among populations (Additional file 2: Table S1).

Table 3 Inference of positively selected amino acid sites for golden pheasant MHC class I sequences

Locus	M8 vs. M7 likelihood ratio test			Positively selected sites			
	df	Test statistic	Significance	OmegaMap	PAML	REL	FEL
<i>IA1</i>							
-E2	2	6.7	<0.01	24 L *	24 L **	24 L **	-
-E3	2	10.5	<0.001	(9 ^A)	99Y**, 155R**	-	-
<i>IA2</i>							
-E2	2	105.8	<0.001	(17 ^B)	9E**, 24 L**, 68 V**, 71 T**, 74I**, 78D**	9E**, 24 L**, 70G**, 71 T**, 78D**	9E**, 24 L**, 71 T**, 74I**, 78D**
-E3	2	47.3	<0.001	(7 ^C)	95 L**, 97 W**, 99Y**, 113H**, 115 V**, 152Y*, 155R**, 156Q*	113H**, 155R**	113H**, 155R*

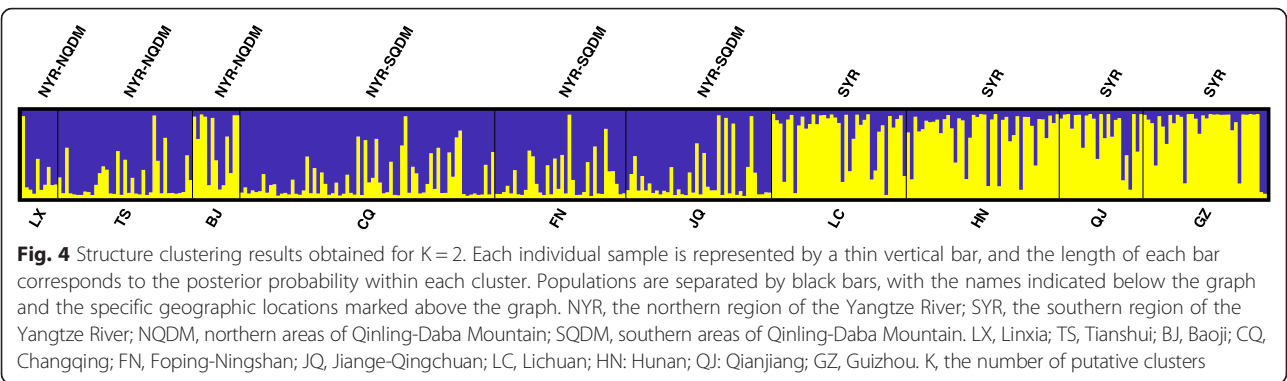
Notes: The test statistic was computed as 2 (Lb-La), where La and Lb are log-likelihood values for the M8 and M7 site model, respectively. *, posterior probability > 0.95; **, posterior probability > 0.99. df, degree of freedom; PAML, phylogenetic analysis by maximum likelihood; REL, random effects likelihood; and FEL, fixed-effects likelihood. ^A The inferred positively selected amino acid sites by OmegaMap in *IA1*-E3 were: 99Y*, 150P**, 151 T**, 152Y**, 153 V**, 154D**, 155R**, 156Q** and 158 L**. ^B The inferred positively selected amino acid sites by OmegaMap in *IA2*-E2 were: 9H**, 24 L**, 40 T*, 41R*, 42R*, 55A**, 56 M*, 57D**, 58Q**, 59Q*, 67I**, 68 V**, 70G**, 71 T**, 74I**, 75Y**, and 78D**. ^C The inferred positively selected amino acid sites by OmegaMap in *IA2*-E2 were: 113H**, 115 V*, 150P**, 151 T*, 152Y**, 155R**, and 156Q**



For *IA1*, most populations exhibited less heterozygosity than expected, and all instances conformed to the Hardy–Weinberg equilibrium (HWE) after Bonferroni correction (Additional file 3: Table S2). For *IA2*, a similar pattern was observed except for 3 cases with significant heterozygote deficits (*IA2-E2* in Hunan [HN] and *IA2-E3* in Changqing [CQ] and Lichuan [LC]; Additional file 3: Table S2). We inferred that the presence of null alleles may not be the main cause of HWE departures in these 3 cases because their frequencies were all <0.1, as estimated by the expectation-maximization algorithm.

The genetic clustering analysis, computed with a non-linkage model in STRUCTURE, revealed that ΔK reached a peak at $K = 2$ (mean values: $\text{LnP}(D) = -5272.2$;

Additional file 4: Figure S2), indicating that the split in the dataset divided the populations into 2 major genetic clusters. The blue cluster (Fig. 4) is dominantly composed of populations scattered across the northern region of the Yangtze River (NYR; including Linxia [LX], Tianshui [TS], CQ, Foping-Ningshan [FN], and Jiange-Qingchuan [JQ]), except for Baoji (BJ). The yellow cluster mainly represents populations located in the southern region of the Yangtze River (SYR; consisting of LC, Qianjiang [QJ], HN, and Guizhou [GZ]; Fig. 4). Based on clustering results, 10 populations were separated into 2 groups for subsequent analysis: NYR (LX, BJ, TS, CQ, FN, and JQ) and SYR (LC, QJ, HN, and GZ). BJ appeared slightly distant from the other NYR populations, probably because of the lack of representative samples. We also performed analysis with a



linkage model in STRUCTURE software. The same ΔK and grouping were obtained, but the relative proportion of each population in either cluster was different. The association between the SYR and NYR groups was clearly demonstrated in both non-linkage and linkage models.

A congruent pattern from STRUCTURE analysis was also obtained when genetic differentiation among populations over the MHC class I loci was estimated by F_{ST} (Additional file 5: Table S3). After omitting the BJ population, pairwise F_{ST} values exhibited no significant differentiation within the NYR or SYR populations (NYR: F_{ST} range 0.001–0.030, $P > 0.01$, mean $F_{ST} = 0.016$; SYR: F_{ST} range 0.009–0.059, $P > 0.01$, mean $F_{ST} = 0.036$), whereas pairwise F_{ST} statistics showed significant differentiation between NYR and SYR populations (F_{ST} range 0.019–0.129, $P < 0.01$; mean $F_{ST} = 0.069$), suggesting that all populations, except BJ, tended to form 2 major groups (NYR [LX, TS, CQ, FN, JQ] and SYR [LC, QJ, HN, GZ]) that were divided by the Yangtze River. A Mantel test showed that the observed correlation between pairwise F_{ST} values and geographical distances was significantly different from zero ($r = 0.389$, $P = 0.013 < 0.05$), indicating that a strong isolation-by-distance relationship occurred for MHC class I variations.

During analysis of molecular variance (AMOVA), we examined the genetic structure pattern of the 2 groups, after completely excluding BJ. The 2-group analysis indicated that 93.65 % of the genetic variation was distributed within populations, 4.07 % of the total variance was partitioned between groups, and 2.28 % occurred among populations within groups (Table 4). The results demonstrated that the genetic variation between groups (V_a) was significantly greater than the genetic variation

among populations (V_b) (Table 4), suggesting that the Yangtze River might serve as a barrier impeding gene flow between the NYR and SYR populations.

Discussion

Extensive allelic variation in MHC class I genes

We characterized, for the first time, the second and third exons of MHC class I genes in 12 wild populations of golden pheasant. We identified alleles for the *IA1* locus (14 E2 alleles and 11 E3 alleles) and for the *IA2* locus (27 E2 alleles and 28 E3 alleles), which differed markedly in their levels of allelic divergence. *IA2* displayed more extensive sequence variations, perhaps because of differential selection pressure or as a reflection of its inherently different properties. Forty-one different alleles in the second exon and 39 different alleles in the third exon were found among 339 individuals, revealing a high polymorphism of 2 peptide-binding domains in the golden pheasant class I genes. In previous studies, other non-model avian species also exhibited extraordinary diversity at exon 3 of MHC class I genes. For example, 36 MHC class I alleles were identified in a screening of 8 red knots [39], 50 unique functional class I variants were detected in a natural population of blue tits [40], and up to 82 divergent class I gene sequences were observed in the scarlet rosefinch population [6]. However, class I sequences from these avian species were generally amplified using species- or motif-specific PCR, which does not involve assignment to particular loci. Hence, for these studies, it was not possible to explore the evolutionary processes underlying MHC variation at each locus.

Table 4 Summary of AMOVA results for MHC class I loci from golden pheasant populations

Source of variation	df	SS	Variance components	Percentage of variation	Fixation index	P
Between groups	1	26.77	V_a : 0.07	4.07	F_{ST} 0.041	<0.00001
Among populations within groups	7	30.08	V_b : 0.04	2.28	F_{SC} 0.024	<0.00001
Within populations	649	1030.58	V_c : 1.59	93.65	F_{ST} 0.064	<0.00001

Notes: Significance tests are based on 1,023 permutations. AMOVA, analysis of molecular variance; df, degree of freedom; SS, sum of squares of deviations from mean

Differential MHC polymorphism driven by differential selection pressure

The main force generating and maintaining MHC polymorphism appears to be balancing selection [41, 42]. Nucleotide sites under positive selection are expected to accumulate more non-synonymous than synonymous substitutions, eventually bringing about amino acid changes and corresponding functional changes in MHC proteins [43]. Such adaptive evolutionary processes due to parasite-driven selection should be evident at the PBR [44]. According to the proposed criteria, both *IA* loci of the golden pheasant showed evidence of positive selection for diversification. For *IA2*-E2, positive selection acted only on PBR codons, which showed a significantly increased degree of non-synonymous substitutions ($d_N/d_S = 9.44$, $P < 0.001$). *IA1*-E2 alleles also evolved under positive selection, although not significantly. E2 alleles of both *IA* genes exhibited higher d_N/d_S ratios at PBR codons than at non-PBR codons, which agrees with the hallmark of such functionally important regions as targets of balancing selection [45]. However, compared with *IA1*, the *IA2* gene showed more allele polymorphisms and more frequent nucleotide substitutions, implying that increased selection pressure more effectively increased the functional variation at *IA2*. Though no significant recombination events were detected in E3 sequences, interlocus homogenization implied that these 2 loci might have evolved by concerted evolution. Previous findings have confirmed that different patterns of allele polymorphisms occurring at MHC genes are subject to the divergence of selection intensities between loci. In humans, the intensity of selection appears to have differed among 3 *HLA* class I loci; *HLA-B*, which was affected by the strongest selection, is more polymorphic than the other 2 class I loci, *HLA-A* and *HLA-C* [46]. In the alpine newt, increased variation in the *DAB* gene, composed of 37 alleles, was largely caused by strong selection over an evolutionary timescale, while low variation in the *DBB* gene, containing only 3 alleles, was correlated with a lack of selection pressure [47]. In contrast, the lack of increased d_N over d_S in the PBR of E3 alleles indicates that positive selection might be less advantageous in the $\alpha 2$ domain. Stronger selection in the $\alpha 1$ domains than in the $\alpha 2$ domains could be ascribed to structural principles, which crucially govern the important peptide-binding motifs of class I molecules [48]. In addition, without exact knowledge of the crystal structures of class I molecules, the putative golden pheasant PBR deduced from consensus sites of chicken [49, 50], might vary considerably from the actual PBR of golden pheasant class I proteins, especially for the $\alpha 2$ domains. However, d_N in the PBR at exon 3, which was twice that determined for the non-PBR, suggests that selection might also have played a role in maintaining their variation.

Based on the predictive outcomes of 4 codon-based models, we found substantial evidence of balancing selection at individual codons, as well as divergent selective pressure between *IA1* and *IA2*. Seven positively selected sites were identified in *IA2* by at least 2 codon-based models (Fig. 2; Table 3), and all of them corresponded to peptide-recognition sites in chickens [49, 50]. A considerable amount of non-synonymous variation occurred at those specific codons under positive selection (Fig. 2), demonstrating that polymorphisms within the $\alpha 1$ and $\alpha 2$ domains may be of functional importance [4]. Conversely, fewer positively selected sites were detected in *IA1*. Therefore, selective pressure on both *IA* loci differed strikingly, and *IA2* underwent much stronger positive selection, especially in the PBR. In conclusion, differing levels of diversity and differential selection might be indicative of different functional roles for the golden pheasant *IA* loci.

Phylogeny and interlocus recombination

The pattern of MHC evolution in birds appears to be quite different from that described in mammals [7, 51–53]. Mammalian MHC genes independently evolved, as reflected by phylogenetic trees showing that corresponding orthologous sequences from different species grouped more closely than other sequences from the same species [25]. In birds, the lack of orthologous relationships among putatively different MHC loci is common, and such examples of phylogeny-based orthology in MHC genes are limited to pairs of closely related avian species [54–57]. However, the phylogenetic inconsistency in avian species is mainly due to higher frequencies of interlocus recombination, also termed “concerted evolution” [51–53, 58]. Frequent recombination events occurring in MHC multigene families usually produce between-locus homogenization, which obscures the true phylogenetic relationship. Similar to some cases observed in class II genes in Galliformes [35, 57, 59], golden pheasant class I sequences failed to constitute well-supported independent groups in accordance with each gene. Instead, class I sequences appeared to group together and were intermixed (Fig. 3), which is consistent with concerted changes. Thus, the occurrence of intergenic recombination between *IA1* and *IA2* was discernibly responsible for the locus-intermingling pattern, underpinning the diversification of MHC genes in the golden pheasant.

Population structure based on MHC class I variation

The *IA1* and *IA2* genes both exhibited a large variation in allele numbers across sampled populations, i.e., 6–22 alleles were found at *IA2*-E2 in each population (Additional file 2: Table S1), as well as in terms of sample size-corrected allelic richness. Alternatively, a large difference in the number of alleles with a frequency of

≤ 0.05 was found among populations, i.e., 0–17 at *IA2-E2* and 5–19 at *IA3-E3* (Additional file 2: Table S1), indicating that rare alleles have been retained in the golden pheasant populations by frequency-dependent selection [15], a potential mechanism accounting for the extreme polymorphism and long-term persistence of alleles at the MHC. Furthermore, the widespread sharing of class I alleles (i.e., *IA2-E2*03* and **04*) across all populations suggests that these alleles may have predominated in golden pheasant populations in recent times. If the same dominant pathogen widely invades different golden pheasant populations, a class I allele that confers a selective advantage in resistance against this prevalent pathogen is quite likely to be shared among populations.

The comparison of genetic differentiation between populations at various spatial scales also supports the hypothesis of balancing selection, which is generally regarded as the main force shaping MHC genetic diversity [14, 42]. In a relatively small geographic region, where populations experience very similar pathogen-mediated selective regimes, i.e., homogenous selection pressure, balancing selection tends to decrease the levels of between-population variation compared to within-population variation [41, 60, 61]. Therefore, the little differentiation among golden pheasant populations from the NYR or SYR might be a result of homogenous balancing selection. However, the structuring of genetic variation at MHC loci indicated that the Yangtze River itself shaped the population structure of golden pheasants. The Yangtze River, as the largest river in China, is too wide for golden pheasants to cross because they do not participate in long-distance migratory flight. Hence, the Yangtze River, an important natural barrier to dispersal and gene flow for golden pheasant populations, intensified the genetic divergence and eventually caused the formation of 2 distinctive population groups: NYR and SYR. Significant inter-population differentiation between the NYR and SYR clusters probably also resulted from geographic heterogeneity of balancing selection between distant populations, which reportedly occurs because of variation in pathogen communities at a broad geographic scale [60, 62–64].

Conclusion

In the present study, we genotyped MHC class I genes from 12 wild golden pheasant populations by PCR-SSCP, using locus-specific primers. Our work revealed that: 1) 2 MHC class I genes exhibited differential genetic polymorphisms, and much stronger positive selection detected at *IA2* than at *IA1* might account for the more extensive variation of *IA2*; 2) interlocus recombination between 2 *IA* genes, noticeably reflected by the intermingling phylogenetic pattern, is also an important mechanism responsible for the extensive allelic variation

of the *IA2* gene; 3) the pattern of population differentiation implied that homogenous balancing selection might explain why an even distribution of MHC variation was maintained among populations within the NYR or SYR region, while the Yangtze River acted as a barrier to gene flow between NYR and SYR populations, and heterogeneous balancing selection might be an important factor determining the NYR-SYR genetic structure in golden pheasants.

Methods

Ethics statement

Muscle tissues were collected from deceased pheasants obtained from poachers and pheasants that died from natural causes, found at nature reserves (NRs). No pheasants were killed specifically for this study. Blood samples were obtained under standard veterinary care during physical examination of rescued birds, i.e., blood samples were not taken solely for the purpose of this study. We collected muscle and blood samples as gene resources with permission from the Department of Wildlife Conservation and Nature Reserve Management under the State Forestry Administration of China, and deposited them (Chpi0001–Chpi0339) in the State Conservation Center for Gene Resources of Endangered Wildlife of China. The pheasant samples used in this study were from LX Taizishan National NR, TS Maicao-gou NR, BJ Shenshahe NR, CQ National NR, FP National NR, NS NR, JG Xihe NR, QC Tangjiahe National NR, LC Xingdoushan National NR, Sangzhi (SZ) Badagongshan National NR, Yongshun (YS) Xiaoxi National NR, Jishou (JS) Baxianhu NR, QJ Wulingshan NR, Jiangkou (JK) Fanjingshan NR, Shibing (SB) Fodingshan NR, and Libo (LB) Maolan National NR. We did not collect samples from any other species at these NRs.

Sample collections from Chinese golden pheasants

We collected 339 golden pheasant muscle and blood samples from the above-mentioned 16 populations throughout Mainland China, of which few samples were collected from pheasants in SZ, YS, and JS (HN Province) or in JK, SB, and LB (GZ Province); thus, the intra-province samples were incorporated into the HN and GZ populations, respectively. Among these 12 populations, 8 populations were distributed within the NYR (3 were scattered among the northern range of the QDM [NQDM]), and the other 5 were located near the southern foothills of the QDM [SQDM]), whereas 4 populations were distributed within the SYR (Fig. 1). Sample sizes of the NS and QC populations were relatively small (7 samples from the NS region and 8 from the QC region); therefore, for convenience in discussing the population structure, we integrated the NS population with the adjacent FP population (i.e., designated as FP-NS or

FN) and the QC population with JG (i.e., designated as JG-QC or JQ). Sample collection proceeded as follows: blood (approximately 100 μ L for each living pheasant) was collected by brachial venipuncture, while muscle tissues were obtained from dead pheasants. Genomic DNA was phenol-chloroform extracted, as described previously [65].

Locus-specific amplifications

Based on direct evidence of cDNA expression in one pheasant (our unpublished experimental data), the 2 MHC class I genes (*IA1* and *IA2*) are expressed. To comprehensively characterize variation in the highly variable regions (exons 2 and 3) in *IA1* and *IA2*, we developed locus-specific screening methods based on knowledge of the MHC genomic structure in the golden pheasant [37]. The locus-specific primer set (A-E2-up1/A-E2-dn1) was designed to amplify fragments of 346 bp in length, which encompassed most of exon 2 in *IA1*. A 10- μ L polymerase chain reaction (PCR) mixture was prepared, which included 20 ng of extracted DNA, 5 μ L of 2 \times GC I Buffer with $MgCl_2$ (TaKaRa, China), 0.2 μ M primers (synthesized by Invitrogen, USA), 0.2 mM dNTPs (TaKaRa, China), and 0.25 U Ex-Taq Polymerase (TaKaRa). The PCR conditions were as follows: 95 $^{\circ}$ C for 5 min; 30 cycles of 95 $^{\circ}$ C for 30 s, 55 $^{\circ}$ C for 30 s, and 72 $^{\circ}$ C for 30 s; followed by 72 $^{\circ}$ C for 10 min. Nested PCR was used to specifically amplify exon 3 of the *IA1* gene. A pair of primers (A-E3-e2up1/A-E3-i3dn1) was used during the first round of amplification, and then a pair of internal primers (A-E3-e2up1/A-E3-e3dn2) was used to amplify a shorter target fragment (272 bp), which was appropriate for subsequent genotype screening. Exon 3 of *IA1* was amplified by nested PCR based on the reaction scheme used to amplify exon 2, except that the cycle number was decreased to 20 in the first round of amplification.

The complete *IA2* gene was amplified with the primers TAP1-E11-upper22 and C4-E1-down11, which targeted long fragments of approximately 4.2 kb spanning from the *TAP1* gene to the *C4* genes [37]. PCR was performed using LA-Taq Polymerase (TaKaRa), and the thermocycling conditions were as follows: an initial step of 94 $^{\circ}$ C denaturation for 1 min; 18 cycles of 98 $^{\circ}$ C for 10 s and 68 $^{\circ}$ C for 5 min; and 72 $^{\circ}$ C for 10 min. Then, using the *IA2* amplicons as templates, 2 other pairs of primers (*Ch.pi*-I1up3/*Ch.pi*-I2dn1 and *Ch.pi*-I2up5/A-E3-e3dn2) were employed in nested-PCRs to amplify exons 2 or 3 of *IA2*, using similar reaction conditions and protocols as described for *IA1*. Consequently, the final products contained only the *IA2* locus. The positions and sequences of all primers used in this study are shown in Figure S3 of Additional file 6.

SSCP genotyping and allele nomenclature

SSCP is a highly sensitive, rapid, convenient, and relatively inexpensive technique for genotyping. By discriminating subtle differences in shifted SSCP bands among different DNA sequences, this method easily detects 1- or 2-nucleotide mutations [66, 67]. Hence, we combined SSCP with DNA sequencing to survey for MHC variations, an approach that has been broadly applied for studying various species [5, 31, 68].

SSCP was performed using a DCode Universal Mutation Detection System (Bio-Rad, USA), according to the manufacturer's instructions and the detailed description by Wan QH et al. [68]. Briefly, 5 μ L of each PCR product was diluted 1:1 in loading dye. The mixture was denatured for 5 min, immediately placed on ice to halt further reaction, and loaded onto a non-denaturing polyacrylamide gel (8–12 % acrylamide). Electrophoresis was conducted at a constant running temperature of 8 $^{\circ}$ C and at 30 W for 8–12 h. SSCP bands were visualized by silver staining [69]. PCR product bands were excised from the gels, and allele sequences were determined by directly sequencing the PCR products. TA cloning and sequencing were also performed to identify novel MHC alleles. We purified the source PCR products using the AxyPrep™ DNA Gel Extraction Kit (Axygen Biosciences, USA), inserted them into the pMD18-T vector (TaKaRa), and transformed the recombinant plasmids into DH5 α competent cells (TaKaRa). In general, at least 3 positive clones per allele were picked for sequencing to confirm their authenticity. To avoid PCR artifacts, all alleles identified in this study were verified at least in 2 different pheasant samples when possible, and alleles found only in one pheasant were validated by performing 3 independent PCRs.

Each single allele was assigned an individual name, consisting of the specific locus (*IA1* or *IA2*), the specific exon (e.g., E2, exon 2), and a sequential number following an asterisk (starting from 01), referring to the nomenclature used for chicken MHC [54]. All *IA* alleles can be accessed in GenBank under the following Accession Numbers: *Chpi-IA1-E2*01–*14* (KM005661–89), *Chpi-IA1-E3*01–*11* (KM005650–60), *Chpi-IA2-E2*01–*27* (KM005703–29), *Chpi-IA2-E3*01–*28* (KM005675–702).

Sequence analysis

Nucleotide sequence alignments and the corresponding amino acid translations were performed using MEGA5 [70]. Estimations of mean pairwise nucleotide distances and amino acid distances were determined by MEGA5. We then employed the modified Nei–Gojobori algorithm [71] via the Jukes–Cantor correction [72] using MEGA5 to compute d_N and d_S . Standard errors were determined as bootstrap values with 1,000 replicates.

The d_N/d_S ratios were evaluated for PBRs, non-PBRs, and all sites of the second and third exons in both investigated genes. Significances were determined by a 1-tailed Z test in MEGA5. The putative peptide-binding residues were inferred from codons corresponding to PBRs of chicken MHC class I molecules [49, 50]. The comparison of codon-based substitution models of sequence evolution for the *IA1* and *IA2* genes was accomplished by 2 methods. First, selections were evaluated using OmegaMap [73]. Without prior knowledge, all codons were assumed to have an equal equilibrium frequency and all priors were set as recommended by Wilson & McVean [73], with the exception of ω and ρ models, which were allowed to independently vary across codons. We ran 500,000 Markov chain Monte-Carlo iterations with a thinning interval of 100. Second, a maximum-likelihood approach in PAML 4.2 [74] was also taken. For simplicity, the nested site models (M7 and M8), which assume ω variation ($\omega = d_N/d_S$) with a beta distribution among sites, were used for our study. In the likelihood-ratio test [75], the pairing models were compared with each other. The M8 model with an extra 11 class sites, allowed the identification of positively selective sites from the Bayes empirical Bayes inference [76], and only sites with posterior probabilities higher than 95 % were considered to be under positive selection. However, for recombination sites, the likelihood methods tend to overestimate ω values among sites and might produce high false positive rates for positively selected sites [77]. Therefore, based on non-recombinant sequences only (see below), 2 additional analyses were applied to infer selection with respect to particular codons using a Bayesian approach: REL [54] assumed the substitution rates across sites under a fixed distribution and then made inferences about the ω rate at each site, while FEL directly estimated ω values for individual sites without any assumption. Both methods were implemented by HyPhy software [78] and are available at the Datamonkey web interface (<http://www.datamonkey.org/>) [79].

Phylogenetic reconstructions using golden pheasant MHC I alleles determined herein and other avian sequences available in the National Center for Biotechnology Information (NCBI) were performed separately for exons 2 and 3. The corresponding class I sequence from tuatara (*Sphenodon punctatus*) was used as an outgroup. A Bayesian phylogeny was reconstructed using MrBayes software, version 3.2 [80] with the posterior probabilities inferred from Metropolis-coupled Markov chain Monte Carlo simulations (MCMCMCs). Two independent MCMCMC simulations (4 chains each, temperatures set at 0.20) were run for 10,000,000 generations with tree sampling every 1,000 generations for both genes until reaching convergence. The first 25,000 sampled trees were excluded as burn-in. The remaining trees were

used to construct the 50 % consensus tree using the GTR (for nucleotide) and JTT (amino acid) models, and the resulting trees were displayed in TreeView [81]. Neighbor-joining (NJ) trees were generated by MEGA5 [82] for comparison with the Bayesian trees. A bootstrap test was performed to estimate support for the nodes in the tree via 1,000 replicates. In addition, Neighbor-Net networks of golden pheasant alleles, together with class I sequences of Galliformes birds, were established using Kimura's 2-parameter model in Splits Tree 4 [83].

RDR4 software [84] was used with the default setting to detect recombination. Recombination analyses were conducted by at least 3 RDR4-implemented algorithms, including RDP [85], GENECONV [86], and MAXCHI [87]; only recombination events that were well verified by all these methods were considered valid. Moreover, we examined the possibility of recombination using the GARD algorithm [88] from the Datamonkey webserver (<http://www.datamonkey.org/>). The haplotypes between exons 2 and 3, as well as inter-exon recombination events were both computed by PHASE version 2.1.1 (<http://stephenslab.uchicago.edu/software.html#phase>).

MHC allele frequencies, number of alleles, allelic richness, H_O and H_E were calculated using FSTAT software version 2.9.3 [89]. The rarefaction method [90, 91] was utilized to correct differences in sample sizes across populations during the measurement of allelic richness. The extent of deviation from HWE was evaluated by Monte Carlo simulation tests with sequential Bonferroni correction [92] in GenePop 4.0 [93]. Based on multilocus genotype data from the studied populations, we inferred the genetic structure by employing the Bayesian clustering algorithm in STRUCTURE V 2.3.3 [94]. We employed non-linkage and linkage models to infer population structures. Typically, the number of putative clusters (K) was assumed to be no more than the number of actual sampled populations; hence, the K value ranged from 1–10. For each K value, 10 independent runs were performed using default settings to obtain log likelihoods estimated by Ln Prob, mean value of ln likelihood, and variance of ln likelihood [95]. The most probable value of K was eventually determined by the model choice criterion ΔK [95], which was dependent on the second-order change of the log likelihood mean. The outputs from the Bayesian analyses were visualized with DISTRUCT v1.1 [96]. Pairwise F_{ST} as an index of population differentiation, was estimated with the Weir and Cockerham method [97] in Arlequin3.5 [98], and the significance levels of F_{ST} values were evaluated by performing 10,000 permutations [99, 100]. Mantel tests were conducted for the detection of isolation by distances at MHC class I genes. First, geographical distances between populations were measured using Google Earth [101]. Then, the correlation between the

natural logarithm of the geographical distance and $F_{ST}/(1-F_{ST})$ for MHC was tested using a simple mantel test in ZT [102] with 10,000 randomizations. AMOVA tests were performed using Arlequin3.5 software [98] to better partition the genetic variation between groups (Va), among populations (Vb), and within populations (Vc).

Availability of supporting data

The data set supporting the results of this article is available in the LabArchives [DOI: 10.6070/H4RR1W8K].

Additional files

Additional file 1: Figure S1. Network tree of exons 2 (A) and 3 (B) for golden pheasant MHC class I genes. (PDF 4 MB)

Additional file 2: Table S1. Allele frequencies at MHC class I genes across golden pheasant populations. (PDF 611 kb)

Additional file 3: Table S2. Summary of MHC genetic diversity in golden pheasant populations. (PDF 365 kb)

Additional file 4: Figure S2. Results of Bayesian clustering analysis performed in STRUCTURE. (PDF 166 kb)

Additional file 5: Table S3. Pairwise F_{ST} values among populations in MHC class I genes. (PDF 261 kb)

Additional file 6: Figure S3. Locus-specific primers amplifying exons 2 and 3 of golden pheasant MHC class I genes. (PDF 2 MB)

Abbreviations

AMOVA: analysis of molecular variance; BJ: Baoji; BS: bootstrap support; CQ: Changqing; d_N : non-synonymous substitutions; d_S : synonymous substitutions; FEL: fixed effects likelihood; FN: Foping-Ningshan; FP: Foping; F_{ST} : F-statistic; GTR: general time reversible; GZ: Guizhou; H_E : expected heterozygosity; HN: Hunan; H_O : observed heterozygosity; HWE: Hardy-Weinberg equilibrium; JG: Jiange; JK: Jiangkou; JQ: Jiange-Qingchuan; JS: Jishou; JTT: jones-taylor-thornton; LB: Libo; LC: Lichuan; LX: Linxia; MCMCMC: metropolis-coupled Markov chain Monte Carlo simulations; MHC: major histocompatibility complex; NCBI: national center for biotechnology information; NJ: neighbor-joining; NQDM: Northern range of Qinling-Daba Mountain; NS: Ningshan; NYR: the northern region of the Yangtze River; PBR: peptide-binding region; QC: Qingchuan; QDM: Qinling-Daba mountain; QJ: Qianjiang; REL: random effects likelihood; SB: Shibing; SQDM: Southern foothill of Qinling-Daba Mountain; SSCP: single-stranded conformation polymorphism; SYR: the southern region of the Yangtze River; SZ: Sangzhi; TS: Tianshui; YS: Yongshun.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

QHW conceived and designed the research; YFG and SGF collected the samples; QQZ, DDS and MMY performed the experiments; QQZ, KH and QHW analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by a grant from the National Natural Science Foundation of China (No. 31070334), a special grant from the State Forestry Administration, and the Fundamental Research Funds for the Central Universities of the P. R. China.

Author details

¹The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, State Conservation Center for Gene Resources of Endangered Wildlife, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China. ²College of Animal Science and Technology, Zhejiang A&F University, Lin'an, Zhejiang 311300, China.

Received: 13 December 2015 Accepted: 2 February 2016

Published online: 18 February 2016

References

- Klein J, Figueroa F. Evolution of the major histocompatibility complex. *Crit Rev Immunol*. 1986;6(4):295–386.
- Natarajan K, Li H, Mariuzza RA, Margulies DH. MHC class I molecules, structure and function. *Rev Immunogenet*. 1999;1(1):32–46.
- Parham P, Ohta T. Population Biology of Antigen Presentation by MHC Class I Molecules. *Science*. 1996;272(5258):67–74.
- Kuduk K, Babik W, Bojarska K, Sliwinski E, Kindberg J, Taberlet P, et al. Evolution of major histocompatibility complex class I and class II genes in the brown bear. *BMC Evol Biol*. 2012;12(1):197.
- Zhu Y, Wan Q-H, Yu B, Ge Y-F, Fang S. Patterns of genetic differentiation at MHC class I genes and microsatellites identify conservation units in the giant panda. *BMC Evol Biol*. 2013;13(1):227.
- Promerová M, Albrecht T, Bryja J. Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*. 2009;61(6):451–61.
- Spurgin LG, Van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC, et al. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Mol Ecol*. 2011;20(24):5213–25.
- Jaratlidsiri W, Isberg S, Higgins D, Gongora J. MHC class I of saltwater crocodiles (*Crocodylus porosus*): polymorphism and balancing selection. *Immunogenetics*. 2012;64(11):825–38.
- Miller HC, Andrews-Cookson M, Daugherty CH. Two Patterns of Variation among MHC Class I Loci in Tuatara (*Sphenodon punctatus*). *J Hered*. 2007;98(7):666–77.
- Kiemnec-Tyburczy K, Richmond J, Savage A, Lips K, Zamudio K. Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. *Heredity*. 2012;109(3):146–55.
- Zhao M, Wang Y, Shen H, Li C, Chen C, Luo Z, et al. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. *BMC Evol Biol*. 2013;13(1):113.
- Evans M, Neff B, Heath D. MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity*. 2010; 104(5):449–59.
- McClelland EK, Ming TJ, Tabata A, Miller KM. Sequence analysis of MHC class I $\alpha 2$ from sockeye salmon (*Oncorhynchus nerka*). *Fish Shellfish Immunol*. 2011;31(3):507–10.
- Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc Biol Sci*. 2010;277(1684):979–88.
- Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*. 1990;124(4):967–78.
- Sinervo B, Calsbeek R. The developmental, physiological, neural, and genetical causes and consequences of frequency-dependent selection in the wild. *Annu Rev Ecol Evol Syst*. 2006;37:581–610.
- Charbonnel N, Pemberton J. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*. 2005;95(5):377–88.
- Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*. 1975;256(5512):50–2.
- Baratti M, DESSI-FULGHERI F, Ambrosini R, BONISOLI-ALQUATI A, Caprioli M, Goti E, et al. MHC genotype predicts mate choice in the ring-necked pheasant *Phasianus colchicus*. *J Evol Biol*. 2012;25(8):1531–42.
- Huchard E, Banié A, Schliehe-Diecks S, Kappeler PM. MHC-disassortative mate choice and inbreeding avoidance in a solitary primate. *Mol Ecol*. 2013;22(15):4071–86.
- Reusch TB, HaEberli MA, Aeschlimann PB, Milinski M. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*. 2001;414(6861):300–2.
- Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H, et al. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol*. 2010;8(1):29.
- Edwards SV, Nussler J, Gasper J. Characterization and evolution of major histocompatibility complex (MHC) genes in non-model organisms, with examples from birds. *Molecular methods in ecology*. 2000;6:168.

24. Mehta RB, Nonaka MI, Nonaka M. Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus *Oryzias*. *Immunogenetics*. 2009;61(5):385–99.
25. 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 U S A*. 1997;94(15):7799–806.
26. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Küstner A, et al. The genome of a songbird. *Nature*. 2010;464(7289):757–62.
27. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 2005;39:121.
28. Nei M, Hughes AL. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: TSUJI K, AIZAWA M, SASAZUKI T, editors. 11th Histocompatibility Workshop and Conference: 6–13 Nov. 1991; Yokohama. New York: Oxford University Press; 1992.
29. Liao D. Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet*. 1999;64(1):24–30.
30. Canal D, Alcaide M, Anmarkrud JA, Potti J. Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Res Notes*. 2010;3(1):236.
31. Hawley DM, Fleischer RC. Contrasting epidemic histories reveal pathogen-mediated balancing selection on class II MHC diversity in a wild songbird. *PLoS One*. 2012;7(1):e30222.
32. Promerová 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(2):285–92.
33. Kaufman J, Jacob J, Shaw J, Walker B, Milne S, Beck S, et al. Gene organisation determines evolution of function in the chicken MHC. *Immunol Rev*. 1999;167(1):101–17.
34. Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken B locus is a minimal essential major histocompatibility complex. *Nature*. 1999;401(6756):923–5.
35. Worley K, Gillingham M, Jensen P, Kennedy L, Pizzari T, Kaufman J, et al. Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. *Immunogenetics*. 2008;60(5):233–47.
36. Zhao ZJ. A Handbook of the Birds of China (Volume I: Non-passerines). Jilin City: Jilin Science and Technology Press; 2001.
37. Ye Q, He K, Wu SY, Wan QH. Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. *PLoS One*. 2012;7(3):e32154.
38. Livant E, Brigati J, Ewald S. Diversity and locus specificity of chicken MHC B class I sequences. *Annu Rev Genet*. 2004;35(1):18–27.
39. Buehler DM, Verkuil YI, Tavares ES, Baker AJ. Characterization of MHC class I in a long-distance migrant shorebird suggests multiple transcribed genes and intergenic recombination. *Immunogenetics*. 2013;65(3):211–25.
40. Wutzler R, Foerster K, Kempenaers B. MHC class I variation in a natural blue tit population (*Cyanistes caeruleus*). *Genetica*. 2012;140(7–9):349–64.
41. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol*. 2003;16(3):363–77.
42. Hedrick PW. Balancing selection and MHC. *Genetica*. 1998;104(3):207–14.
43. Hughes ALYM. Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet*. 1998;32(1):415–35.
44. Potts WK, Wakeland EK. Evolution of diversity at the major histocompatibility complex. *Trends Ecol Evol*. 1990;5(6):181–7.
45. Satta Y, Li YJ, Takahata N. The neutral theory and natural selection in the HLA region. *Front Biosci*. 1998;3:459–67.
46. Satta Y, O'Huigin C, Takahata N, Klein J. Intensity of natural selection at the major histocompatibility complex loci. *Proc Natl Acad Sci U S A*. 1994;91(15):7184–8.
47. Babik W, Pabijan M, Radwan J. Contrasting patterns of variation in MHC loci in the Alpine newt. *Mol Ecol*. 2008;17(10):2339–55.
48. Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *J Mol Evol*. 1998;281(5):929–47.
49. Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny HJ, et al. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity*. 2007;27(6):885–99.
50. Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc Natl Acad Sci U S A*. 2006;103(5):1434–9.
51. Alcaide M, Edwards SV, Negro JJ. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *J Mol Evol*. 2007;65(5):541–54.
52. Edwards SV, Wakeland E, Potts W. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. *Proc Natl Acad Sci U S A*. 1995;92(26):12200–4.
53. Hess CM, Edwards SV. The Evolution of the Major Histocompatibility Complex in Birds Scaling up and taking a genomic approach to the major histocompatibility complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. *Bioscience*. 2002;52(5):423–31.
54. Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, et al. 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics*. 2004;56(4):261–79.
55. Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L. Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Mol Biol Evol*. 2008;25(6):1180–91.
56. Strand T, Westerdahl H, Höglund J, Alatalo RV, Siitari H. The Mhc class II of the Black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. *Immunogenetics*. 2007;59(9):725–34.
57. Wittzell H, Bernot A, Auffray C, Zoorob R. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Mol Biol Evol*. 1999;16(4):479–90.
58. Bonneaud C, Sorci G, Morin V, Westerdahl H, Zoorob R, Wittzell H. Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics*. 2004;55(12):855–65.
59. Eimes JA, Bollmer JL, Dunn PO, Whittingham LA, Wimpee C. Mhc class II diversity and balancing selection in greater prairie-chickens. *Genetica*. 2010;138(2):265–71.
60. Ekblom R, Saether SA, Jacobsson P, Fiske P, Sahlman T, Grahm M, et al. Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Mol Ecol*. 2007;16(7):1439–51.
61. Miller K, Kaukinen K, Beacham T, Withler R. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica*. 2001;111(1–3):237–57.
62. Cammen K, Hoffman J, Knapp L, Harwood J, Amos W. Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Mol Ecol*. 2011;20(4):740–52.
63. Loiseau C, Richard M, Garnier S, Chastel O, Julliard R, Zoorob R, et al. Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Mol Ecol*. 2009;18(7):1331–40.
64. Sonsthagen SA, Fales K, Jay CV, Sage GK, Talbot SL. Spatial variation and low diversity in the major histocompatibility complex in walrus (*Odobenus rosmarus*). *Polar Biol*. 2014;37(4):497–506.
65. Sambrook J, Russell DW. Molecular cloning: a laboratory manual. third. New York: Cold Spring Harbor Laboratory Press; 2001.
66. Hayashi K. PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *Genome Res*. 1991;1(1):34–8.
67. Sunnucks P, Wilson A, Beheregaray L, Zenger K, French J, Taylor A. SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol*. 2000;9(11):1699–710.
68. Wan QH, Zhang P, Ni XW, Wu HL, Chen YY, Kuang YY, et al. A novel HURRAH protocol reveals high numbers of monomorphic MHC class II loci and two asymmetric multi-locus haplotypes in the Père David's deer. *PLoS One*. 2011;6(1):e14518.
69. Beidler JL, Hilliard PR, Rill RL. Ultrasensitive staining of nucleic acids with silver. *Anal Biochem*. 1982;126(2):374–80.
70. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–9.
71. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3(5):418–26.

72. Jukes T, Cantor C. Evolution of protein molecules. *Mamm Protein Metab.* 1969;3:21–132.
73. Wilson DJ, McVean G. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics.* 2006;172(3):1411–25.
74. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
75. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* 1998;148(3):929–36.
76. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005;22(12):2472–9.
77. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006;55(4):539–52.
78. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;21:676–9.
79. Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21(10):2531–3.
80. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42.
81. Page RD. TREEVIEW, tree drawing software for Apple Macintosh and Microsoft Windows. *Comput Appl Biosci.* 1996;12:357–8.
82. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8(3):275–82.
83. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23(2):254–67.
84. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuve P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 2010;26(19):2462–3.
85. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics.* 2000;16(6):562–3.
86. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology.* 1999;265(2):218–25.
87. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol.* 1992;34(2):126–9.
88. Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. *Bioinformatics.* 2006;22(24):3096–8.
89. Goudet J. FSTAT: A Program to Estimate and Test Gene Diversities and Fixation Indices, Version 2.9.3.2. 2002. Available at <http://www.unil.ch/Jahia/site/dee/op/edit/pid/36921>.
90. Hurlbert SH. The nonconcept of species diversity: a critique and alternative parameters. *Ecology.* 1971;52(4):577–87.
91. Petit RJ, El Mousadik A, O P. Identifying populations for conservation on the basis of genetic markers. *Conserv Biol.* 1998;12(4):844–55.
92. Rice WR. Analyzing tables of statistical tests. *Evolution.* 1989;223–225.
93. Roussel F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour.* 2008;8(1):103–6.
94. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
95. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14(8):2611–20.
96. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes.* 2004;4(1):137–8.
97. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution.* 1984;38(6):1358–70.
98. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564–7.
99. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: chi 2 and the problem of small samples. *Mol Biol Evol.* 1989;6(5):539–45.
100. Roff DA, Bentzen P. Detecting geographic subdivision: A comment on a paper by Hudson et al. *Mol Biol Evol.* 1992;9:968.
101. Google Earth. <http://earth.google.com>.
102. Bonnet E, Van de Peer Y. ZT: a software tool for simple and partial Mantel tests. *Stat Softw.* 2002;7(10):1–12.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

