Immunogenetics (2019) 71:49–59
https://doi.org/10.1007/s00251-018-1076-0

ORIGINAL ARTICLE

Polymorphism of duck MHC class molecules

Lin Zhang[1] · Dongmei Lin[2] · Sen Yu[1] · Junping Bai[1,2] · Wanchun Jiang[2] · Wenzheng Su[1] · Yanyan Huang[1] · Shaohua Yang[1] · Jiaqiang Wu[1]

Received: 25 April 2018 / Accepted: 10 August 2018 / Published online: 5 September 2018

Abstract

Major histocompatibility complex class I (MHC I) molecules are critically involved in defense against pathogens, and their high polymorphism is advantageous to a range of immune responses, especially in duck displaying biased expression of one *MHC I* gene. Here, we examined MHC I polymorphism in two duck (*Anas platyrhynchos*) breeds from China: Shaoxing (SX) and Jinding (JD). Twenty-seven unique UAA alleles identified from the *MHC I* genes of these breeds were analyzed concerning amino acid composition, homology, and phylogenetic relationships. Based on amino acid sequence homology, allelic groups of *Anas platyrhynchos* MHC I (*Anpl*-MHC I) were established and their distribution was analyzed. Then, highly variable sites (HVSs) in peptide-binding domains (PBD) were estimated and located in the three-dimensional structure of *Anpl*-MHC I. The UAA alleles identified showed high polymorphism, based on full-length sequence homology. By adding the alleles found here to known *Anpl*-MHC I genes from domestic ducks, they could be divided into 17 groups and four novel groups were revealed for SX and JD ducks. The UAA alleles of the two breeds were not divergent from the MHC I of other duck breeds, and HVSs were mostly located in the peptide-binding groove (PBG), suggesting that they might determine peptide-binding characteristics and subsequently influence peptide presentation and recognition. The results from the present study enrich *Anpl*-MHC I polymorphism data and clarify the distribution of alleles with different peptide-binding specificities, which might also accelerate effective vaccine development and help control various infections in ducks.

Keywords Duck · Major histocompatibility class I · Polymorphism · Allelic group · Peptide

Introduction

Major histocompatibility complex class I (MHC I) molecules play a critical role in adaptive immune responses by presenting pathogen-derived epitope peptides to specific T cell receptors (TCRs), which result in the proliferation of cytotoxic T lymphocytes (CTLs) and eventual elimination of pathogens from the host (McMichael et al. 1983; Kane and Clark 1984; Jondal et al. 1996). An effective CTL response relies on the ability of MHC I proteins to present a diverse array of peptides. This is achieved through polygenic, polymorphic, and codominant expression of class I genes (Moon et al. 2005; Chan et al. 2016). High MHC I polymorphism, which is mostly clustered at α1 and α2 domains, is thought to confer disease protection at the population level (Bjorkman and Parham 1990; Lie et al. 1991; Chan et al. 2016; Fan et al. 2018). In human and mice, three functional MHC I loci provide polymorphic alleles for presenting peptides (Klein et al. 1983; Rawle et al. 1991; McKenzie et al. 1999), and this high degree of diversity is considered to be associated with resistance to numerous infectious diseases (Zhang et al. 2004; Takeshima et al. 2009).

Several avian species, such as chicken and duck, predominantly express a single *MHC I* gene (Kaufman 1999; Moon et al. 2005; Chan et al. 2016). In chicken, BF2, which is adjacent to *transporter associated with antigen processing* (*TAP*) gene, is the only dominant MHC I locus expressed (Kaufman et al. 1999; Shaw et al. 2007). Although there are five different loci (UAA, UBA, UCA, UDA, and UEA) in the MHC I genome

Lin Zhang and Dongmei Lin contributed equally to this work.

✉ Lin Zhang
lzhang3406@163.com

✉ Jiaqiang Wu
443794629@qq.com

[1] Shandong Key Laboratory of Disease Control and Breeding, Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Science, No. 8, Sangyuan Road, Jinan, Shandong, China

[2] College of Life Sciences and Food Engineering, Hebei University of Engineering, 199, Guangming South Street, Handan, Hebei, China

region of duck (*Anpl*-MHC I), only UAA, which is adjacent to the polymorphic *TAP2* gene, is highly expressed (Mesa et al. 2004; Moon et al. 2005; Chan et al. 2016; Fleming-Canepa et al. 2016). Chicken MHC has an obvious characteristic that associated with resistance and susceptibility to pathogens. Over decades of experiments, it has been found that some MHC haplotypes (such as B21) generally confer resistance to a variety of infectious viral disease, whereas other haplotypes (including B4) often confer susceptibility (Bacon et al. 1981; Bacon and Witter 1993; Plachy et al. 1992). These results were closely related to BF, BG, or some other genes (Plachy and Benda 1981; Hepkema et al. 1993; Kaufman et al. 1999; Goto et al. 2009); among these genes, the highly variable polymorphic molecules, BF2, which are coevolving with polymorphic *TAP* genes (Mesa et al. 2004; Walker et al. 2005; Wallny et al. 2006), determine their peptide-binding characteristics and influence adaptive immune response, and, thus, also play an important role in disease resistance and susceptibility (Kaufman et al. 1995; Wallny et al. 2006; Koch et al. 2007; Zhang et al. 2012).

According to our previous study (Liu et al. 2017), and to several published reports (Bingham et al. 2009; Liang et al. 2011; Li et al. 2015), *Anpl*-MHC I also plays a role in susceptibility to viral infections. The high expression of MHC I and the high resistance of some duck lines to avian influenza virus, Newcastle disease virus, or duck Tembusu virus indicated that the MHC I alleles of such lines bind more virus peptide epitopes. However, little is still known about the associations between particular *Anpl*-MHC I alleles and infectious disease, because few *Anpl*-MHC I alleles have been characterized in few duck lines (Moon et al. 2005; Fleming-Canepa et al. 2016; Zhang et al. 2017), and only one *Anpl*-MHC I crystal structure (*Anpl*-UAA*01) has been resolved (Wu et al. 2017). The polymorphism and peptide-binding specificity of *Anpl*-MHC I in other duck breeds remain unknown. Limited studies on *Anpl*-MHC I polymorphism and its functional classification are the primary bottlenecks hampering duck CTL immunity research and the development of duck lines resistant to viruses.

Shaoxing (here named SX) and Jinding (here named JD) are local elite duck breeds in southeast China, and both have a long history of propagation owing to their outstanding fertility and egg-laying performance. In the present study, 27 *Anpl-MHC I* genes located in locus UAA of SX and JD lines were cloned, and all their alleles were analyzed through multiple sequence alignment and phylogenetic tree construction. Then, the coefficient of variance of the PBD region of each allelic group was estimated based on the Wu-Kabat variability index, followed by the location of the highly variable sites (HVSs) on reported crystal structure models. The combined results of sequence data and crystal structure provide valuable insight into the polymorphism and diversity of *Anpl-MHC I* genes, facilitating further studies on disease resistance differences between duck breeds and the development of CTL epitope vaccines to prevent diseases in ducks.

Methods

Animals

The animals used to analyze MHC I polymorphism belonged to two Chinese egg-laying duck lines, including seven SX ducks and seven JD ducks. Ducks from both breeds were captured at 12 weeks of age from large flocks of two different farms which located in Zhejiang and Fujian provinces, respectively, where are native breeding regions of SX and JD ducks in China. Spleen tissues were dissected from 12-week-old ducks and used for RNA extraction.

RNA extraction, cDNA synthesis, and *Anpl-MHC I* gene amplification

Total RNAs of fresh spleen tissue samples from SX and JD ducks were extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Total RNAs (2 μg per sample) were then used to synthesize first-strand cDNAs by reverse transcription with Oligo (dT)18 primers (TaKaRa Biotechnology, Dalian, China). A pair of reported UAA allele group-specific PCR primers (Zhang et al. 2017) was used to amplify the cDNA segments containing the coding domains of UAA. The PCR was carried out in a final volume of 50 μL, including 0.5 μg cDNA template, 100 pmol of each primer, and 0.5 U LA *Taq* polymerase (TaKaRa Biotechnology). Amplification comprised an initial denaturation at 98 °C for 5 min, followed by 32 cycles at 94 °C for 1 min, 65 °C for 1 min, and 72 °C for 2 min, and a final extension at 72 °C for 10 min. Amplicons of about 1.1 kb were separated by agarose gel electrophoresis, purified using the DNA recovery kit (Tiangen Biotech, Beijing, China), and subcloned to the pMD18-T easy vector (TaKaRa Biotechnology) according to the manufacturer's recommendations. The constructs were transformed into competent *Escherichia coli* DH5α (Transgen Biotech Co., Ltd., Beijing, China) coated on lysogeny broth (LB) plates containing ampicillin (100 μg/mL), isopropyl β-D-1-thiogalactopyranoside (40 μg/mL), and X-gal (20 μg/mL). After incubation overnight at 37 °C, 12 white spots on each LB plate were identified by restriction enzyme analysis with *Eco*R I and *Hin*d III (TaKaRa Biotechnology), and positive clones were then sequenced by Sangon Biotech Co., Ltd. (Shanghai, China). A cDNA sequence was considered a real allele when sequences from two different clones from independent PCR reactions or from different individuals were identical. Sequences were submitted to the National Center for Biotechnology Information GenBank (NCBI, https://www.ncbi. nlm.nih.gov/genbank/) and their accession numbers are displayed in Table 2.

Sequence analysis and phylogenetic tree construction

Obtained sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and JALVIEW (Waterhouse et al. 2009). Based on the alignment of the extracellular domain of all *Anpl*-MHC I nucleotide and amino acid sequences, phylogenetic tree was constructed in MEGA6 (Zhou et al. 2004) using the neighbor-joining method and p-distance (Saitou and Nei 1987). Bootstrap support values were obtained from 1000 replicates.

HVS location in three-dimensional UAA structures

The variation coefficients of the α1 and α2 domains of UAA were calculated on the PVS website (http://imed.med.ucm.es/PVS/). Highly variable sites of all sequences were analyzed using the Wu-Kabat method (Kabat et al. 1997). Then, HVSs (Wu-Kabat index > 6.0) were located and labeled on the MHC I three-dimensional structures using PyMOL Molecular Graphics System (http://www.pymol.org/). The crystal structures used for modeling MHC I molecules from JD and SX breeds built using SWISS-MODEL (http://swissmodel.expasy.org). were *Anpl*-UAA*01 complexes (protein data bank (PDB) codes 5GJX and 5GJY).

Accession numbers

The nucleotide sequences constructed the phylogenetic tree were retrieved from NCBI database under the accession numbers HSU18930 (HLA-A2), M58156 (mouse. H-2K-f), AM282699 (BF2*0401), AY234769 (BF2*2101), AB115245 (*Anpl*-UCA01 (Xia et al. 2004), designated crystal/*Anpl*-UAA*01 (Wu et al. 2017) in the present study), AB115241 (*Anpl*-UAA02), AB115244 (*Anpl*-UBA01), AB115243 (*Anpl*-UBA02), AB115246 (*Anpl*-UDA01), AY294416 (*Anpl*-U02), AY294417 (*Anpl*-U03), AY294418 (*Anpl*-U04), AY294419 (*Anpl*-U05), KX118673 (UAA01*WS), KX118675 (UAA03*WS), KX118676 (UAA04*WS), KX118677 (UAA05*WS), KX118679 (UAA07*WS), KX118680 (UAA08*WS), KX118681 (UAA09*WS), KX118683 (UAA01*CV), KX118684 (UAA02*CV), KX118685 (UAA03*CV), KX118686 (UAA04*CV), KX118687 (UAA05*CV), KX118688 (UAA06*CV), and KX118689 (UAA07*CV).

Results

Sequence features

In most individuals, two or three alleles were recovered in many clones (Table 1). Finally, 27 new *Anpl*-MHC I alleles were identified at the UAA locus: 15 from the seven SX ducks (UAA01*SX–UAA05*SX and UAA07*SX–UAA16*SX) and 12 from the seven JD ducks (UAA01*JD–UAA02*JD, UAA04*JD–UAA08*JD, and UAA10*JD–UAA14*JD). Their accession numbers are listed in Table 2. In addition, one unassigned allele, likely UDA repeated four times, was recovered from duck SX-3 (Table 1).

Full-length amino acid sequences corresponding to these alleles were aligned with crystal/*Anpl*-UAA*01 using Clustal Omega and analyzed using JALVIEW. As shown in Fig. 1, 353 or 354 amino acids were coded, including signal peptide and extracellular α1, α2, and α3 domains, followed by transmembrane and cytoplasmic (TM/CY) domains. The signal peptide and α1, α2, and α3 domains comprised 21, 88, 92,

Table 1 Alleles expressed in cDNA in each duck

Duck	UAA alleles	Unassigned alleles
SX-1[a]	UAA01*SX (6)	
	UAA02*SX (4)	
SX-2	*UAA03*SX (5)*	
	UAA04*SX (2)	
	UAA05*SX (4)	
SX-3	UAA07*SX (5)	U17*SX (4)
	*UAA03*SX (2)*	
SX-4	UAA08*SX (4)	
	UAA09*SX (3)	
SX-5	UAA10*SX (2)	
	UAA11*SX (3)	
	UAA12*SX (3)	
SX-6	UAA13*SX (5)	
	UAA14*SX (3)	
SX-7	UAA15*SX (4)	
	UAA16*SX (4)	
JD-1	UAA01*JD (10)	
JD-2	*UAA02*JD (5)*	
JD-3	UAA04*JD (2)	
	*UAA05*JD (4)*	
JD-4	UAA06*JD (3)	
	UAA07*JD (2)	
	*UAA08*JD (3)*	
JD-5	UAA10*JD (3)	
	UAA11*JD (2)	
	*UAA05*JD (4)*	
JD-6	UAA12*JD (2)	
	UAA13*JD (3)	
	*UAA08*JD (2)*	
JD-7	UAA14*JD (6)	
	*UAA02*JD (4)*	
	*UAA03*SX (2)*	

The number of matching clones is indicated in parentheses. Shared alleles are indicated in italics

[a] Individuals in different duck lines

Table 2 Gene groups of *Anpl*-MHC I in Shaoxing (SX) and Jinding (JD) ducks

Sequence name	Allelic groups	Alignment with the known sequences	GenBank accession no.
SX			
UAA01*SX	8		MH218821
UAA02*SX	8		MH218823
UAA03*SX	5		MH218824
UAA04*SX	17	New[a]	MH218826
UAA05*SX	5		MH218828
UAA07*SX	5		MH218831
UAA08*SX	5		MH218833
UAA09*SX	7		MH218834
UAA10*SX	17	New	MH218836
UAA11*SX	17	New	MH218838
UAA12*SX	17	New	MH218840
UAA13*SX	8		MH218842
UAA14*SX	8		MH218844
UAA15*SX	5		MH218845
UAA16*SX	5		MH218846
JD			
UAA01*JD	11		MH218820
UAA02*JD	12		MH218822
UAA04*JD	2		MH218825
UAA05*JD	2		MH218827
UAA06*JD	14	New	MH218829
UAA07*JD	15	New	MH218830
UAA08*JD	5		MH218832
UAA10*JD	2		MH218835
UAA11*JD	11		MH218837
UAA12*JD	16	New	MH218839
UAA13*JD	5		MH218841
UAA14*JD	5		MH218843
White Peking			
Anpl-U02	1		AY294416
Anpl-U03	2		AY294417
Anpl-U04	3		AY294418
Anpl-U05	4		AY294419
Peking[b]			
Anpl-UAA01	5		AB115242
Anpl-UAA02	6		AB115241
Anpl-UBA01	7		AB115244
Anpl-UBA02	8		AB115243
Anpl-UCA01	9		AB115245
Anpl-UDA01	10		AB115246
WS			
UAA01*WS	11		KX118673
UAA03*WS	11		KX118675
UAA04*WS	11		KX118676
UAA05*WS	12		KX118677
UAA07*WS	8		KX118679
UAA08*WS	5		KX118680
UAA09*WS	12		KX118681
CV			
UAA01*CV	8		KX118683
UAA02*CV	1		KX118684
UAA03*CV	1		KX118685
UAA04*CV	1		KX118686
UAA05*CV	1		KX118687
UAA06*CV	8		KX118688
UAA07*CV	13		KX118689

[a] Novel allelic groups identified in this study

[b] *Anpl*-MHC I sequences reported in reference (Xia et al. 2004)

and 91 amino acids, respectively. The length of the TM/CY domain varied in some duck UAA alleles reported previously (Zhang et al. 2017). Among the 27 sequences found here, most TM/CY domains contained 62 amino acids, with only two sequences containing 61 amino acids. Sequence polymorphism was mostly distributed in α1 and α2 domains, which collectively formed the PBG, suggesting that such sequence diversity is functional. Moreover, four cysteine residues (positions 99, 162, 200, and 256) that might form two couple disulfide bonds in α2 and α3 domains and an N-linked glycosylation site (NQS) in the α1 domain were highly conserved. The amino acids involved in peptide anchoring include 7Y, 59Y, 84R, 123Y, 143T, 146K, 147W, 159Y, and 171Y (numbering relative to HLA-A2). These were also conserved in the *Anpl*-MHC I sequences, which included 7Y, 58Y, 83R, 140T, 143K, 144W, 157Y, and 169Y amino acids, with only 123Y being replaced by 120F.

Classical MHC I molecules interact with CD8 molecules, which is essential for T cell activation. In the 27 sequences reported here, amino acids 112Q and 223Q, which were involved in the interaction with CD8 molecules, corresponded to amino acids 114Q and 226Q in mammal MHC I (Kaufman et al. 1992; Fan et al. 2018), thus being highly conserved.

According to the crystal structure of *Anpl*-UAA*01 (Wu et al. 2017), 13 amino acids (18G, 35R, 46R, 94Q, 117K, 118D, 185R, 187S, 200R, 228S, 231P, 234D, and 240W) interacted with β2m. In the 27 *Anpl*-MHC I sequences reported here, four amino acids (35R, 117K, 118D, and 185R) were partially conserved while nine amino acids were highly conserved (Fig. 1).

Gene groups and their distribution

None of the sequences of the 27 alleles identified from SX and JD ducks was identical to that reported for other duck breeds. To investigate homology in *Anpl*-MHC I molecules and their coverage in domestic duck lines, the 27 full-length UAA sequences found here were analyzed with that reported for other duck breeds, such as Peking duck, Cherry Valley (CV) duck, and Weishan (WS) Ma duck. Based on amino acid sequence homology, *Anpl*-MHC I sequences were divided into 17 groups, each containing sequences with similarity over 91% (Yan et al. 2005; Zhang et al. 2017). Additionally, SX and JD ducks contained three and one novel allelic groups, respectively (Table 2). As a reference for functional classification (Zhang et al. 2017), the distribution of gene groups in different duck breeds and individuals was subsequently analyzed. As shown in Table 2, 8 of 17 groups contained two or more unique alleles and almost all alleles in each group were distributed in different duck breeds. In SX and JD breeds, there were 10 allelic groups; 4 (5, 7, 8, and 17) were distributed in SX ducks and 7 (2, 5, 11, 12, 14, 15, and 16) in JD ducks. Except for SX-2 and SX-4 ducks, all UAA alleles expressed in each SX individual belonged to the same allelic group.

However, in JD ducks, alleles isolated from JD-4, JD-5, JD-6, and JD-7 individuals were dispersed in different groups. Moreover, the allele, UAA03*SX, was also isolated from JD-7 (Table 1). These results indicated that different gene groups were distributed in different duck lines and individuals, suggesting that high allelic diversity is perpetuated in ducks. The expression of one MHC class I gene is biased in ducks, which might facilitate viral escape, but high diversity is beneficial to infection resistance (Fleming-Canepa et al. 2016). At both the breed and individual levels, allele polymorphism was higher in JD than in SX ducks, which might be the result of strong environmental selection pressures during their evolution.

Phylogenetic analysis of the identified UAA alleles

To determine the phylogenetic relationships of UAA alleles between the two duck breeds, nucleotide and amino acid phylogenetic trees were constructed based on the sequences of the extracellular domains of all the UAA alleles identified in this study and of *Anpl*-MHC I sequences, including UDA and UEA, retrieved from GenBank. In addition, to examine the phylogenetic relationships with other species, nucleotide and deduced amino acid sequences of HLA-A2, *Mus musculus* MHC I (H-2), and BF2 were also retrieved and aligned. As shown in Fig. 2a, b, all *Anpl*-MHC I sequences clustered together, containing UDA and UEA which had similar gene sequences and hallmarks of classical MHC I genes, while HLA-A2, BF2, and H-2k-f alleles were independent from the duck MHC I cluster. Furthermore, in the *Anpl*-MHC I cluster, the UAA alleles of SX and JD ducks were not segregated from the UAA alleles of other duck breeds, indicating a close phylogenetic relationship and suggesting a concerted evolution of UAA alleles among duck species.

HVS location in the three-dimensional structure of *Anpl*-MHC I

Using the Wu-Kabat method, HVSs with variability scores > 6.0 in the α1 and α2 domains of the 27 sequences were analyzed. As shown in Fig. 3a, there were 13 HVSs located in the α1 domain and 6 HVSs in the α2 domain, most clustering at amino acids 61 to 75. High variability scores (> 10) appeared at amino acid positions P9, P61, P62, P66, P68, P93, and P95, with P9 presenting the highest score (21.7). Analysis at the breed level evidenced 11 HVSs in the two functional domains (8 in α1 and 3 in α2) of the 15 alleles of SX ducks, with high Wu-Kabat scores (> 10) at positions P9, P95, P62, and P111 (Fig. 3b). In JD ducks, HVSs clustered mostly at the α1 domain (11 HVSs), while the α2 domain contained only two HVSs; amino acid positions P9 and P66 showed the highest scores with Wu-Kabat index values > 20 (Fig. 3c).

The comprehensive analysis of HVS variability scores in the two breeds revealed both overlapping and offsetting effects due to amino acid variation (Table 2). Amino acids at positions P61, P62, P93, and P95 acted as HVSs both in the analysis of all 27 sequences or in the analysis performed for each single breed, although showing higher variation in the analysis of the 27 sequences. Another six amino acids, at positions P22, P43, P69, P116, P128, and P154, also presented higher variability scores in the analysis of the 27 sequences than that of each single breed, even when they were not HVSs in one or both breeds. However, not all HVSs displayed a cumulative effect (Table 3). Thirteen amino acids showed only an offsetting effect on HVS variability scores, with lower variation in the analysis of the 27 sequences than in the analysis of each single breed or both breeds. These results suggest that the functional domains of UAA exhibit highly variable polymorphism, contributing to present more antigenic peptides and to display conserved amino acid sequences in different duck lines, despite allelic variation.

Half of the HVSs composing the PBD were located in pockets in the three-dimensional structure of UAA (Fig. 3d). Among them, amino acids at positions P9, P24, P43, P62, P66, and P69 were located in the B pocket, at P95 and P111 in the C pocket, at P111 and P154 in the E pocket, and at P93 in the F pocket. Other HVSs were mainly located in α helix, β strand, and loop regions of the α1 domain, playing an important role in forming the antigen binding cleft. The HVSs of UAA alleles from SX ducks were clustered in the B, F, C, and E pockets, with 50% being located in the B pocket (Fig. 3e), which is vital for peptide presentation (Wu et al. 2017). In the UAA alleles of JD ducks, five of the eight HVSs located in pockets were clustered in the B pocket (Fig. 3f). According to these results, no HVSs were located in the A and D pockets. The exceptional clustering of HVSs might be specific to the two breeds examined here or and effect of the limited sample size. The combined analysis of Wu-Kabat variability plots and location of HVSs in the crystal structure of UAA indicated that most diversity is localized within the PBG, which is important for broadening the spectra of antigenic peptide presentation.

Discussion

To date, only 86 *Anpl-MHC I* genes have been investigated from limited duck lines (Xia et al. 2004; Moon et al. 2005; Fleming-Canepa et al. 2016; Zhang et al. 2017). Although there are 24 indigenous duck breeds in China, SX and JD ducks are representative elite egg-laying breeds. In the present study, 27 unique MHC I alleles from 14 ducks were identified, typically two or three expressed alleles in each individual. The conserved amino acids identified in the MHC I molecules of mammals and other birds were also found in SX and JD ducks

(Fig. 1), including the cysteines involved in the formation of disulfide bonds and the amino acids interacting with β2m and CD8 molecules (Kaufman et al. 1992; Xia et al. 2004; Moon et al. 2005; Wu et al. 2017; Fan et al. 2018). Extensive polymorphism appeared in α1 and α2 domains, which are involved in determining PBG specificity, and the amino acids constituting the α3 domain were relatively conserved. There were more polymorphisms in the α1 domain than in the α2 domain, and this was consisting with the recently resolved crystal structures of *Anpl*-UAA*01 which revealed that almost all amino acids comprising the B and F pockets, which act as primary anchor sites for peptides, were located in the α1 domain (Wu et al. 2017). Thus, higher polymorphism in the α1 domain has a greater influence on binding peptide specificity. The 27 alleles were highly homologous with other duck *Anpl*-MHC I and appeared in the same cluster in the phylogenetic tree. Meanwhile, some identical extracellular domain sequences were found in more than one duck or one breed, such as *Anpl**U02, UAA02*CV, UAA03*CV, UAA04*CV, and UAA05*CV (Fig. 2). These results indicated that there is no significant divergence of *Anpl*-MHC I between Chinese duck breeds. In ducks, there were two expressed *Anpl*-MHC I genes, UAA and UDA. The UAA gene is biased expressed at levels 10-fold greater than that of another gene, UDA, which is the poorly expressed gene (Moon et al. 2005; Chan et al. 2016). In the present study, four cDNA sequences (U17*SX), which were likelihood of UDA alleles, were recovered from SX-3 duck individual, but not from other individuals (Table 1). As the expression level in different tissues was absent, the alleles were temporarily divided into unassigned alleles. By counting the number of UAA and other alleles expressed in SX-3, UAA was present at approximately 1.75-fold abundance of UDA in SX-3, which is lower than that of previous studies (Moon et al. 2005). The lower ration of UAA versus UDA alleles and the less recovery of UDA from duck individuals may be primarily influenced by the limited numbers of available sequences obtained in each duck, or else determined by the individual difference of ducks captured in this study.

Several alleles at the UAA locus contained a unique motif: two additional cysteines (Cys) at positions 95 and 112 in the α2 domain (Fleming-Canepa et al. 2016; Wu et al. 2017), corresponding to positions 97 and 116 in HLA-A2 (Borbulevych et al. 2011). Fleming-Canepa et al. (2016) considered that the Cys95 and Cys112 found in alleles of wild mallards were a conserved structure (Hazes and Dijkstra 1988) but unable to form a disulfide bond. However, in the structure of *Anpl*-UAA*01, an allele from Peking duck, Cys95 and Cys112 form a disulfide bond able to ligate the β5 and β6 sheets at the bottom of PBG to enhance peptide binding (Wu et al. 2017). In the present study, Cys95 and Cys112 were absent from all newly cloned UAA alleles, indicating that this unique motif is not common in *Anpl*-MHC I alleles and that the additional disulfide bond might only be formed in specific individuals. Although our results might have been affected by the limited number of isolated genes, the low genetic coverage of Chinese duck lines is also an indisputable fact.

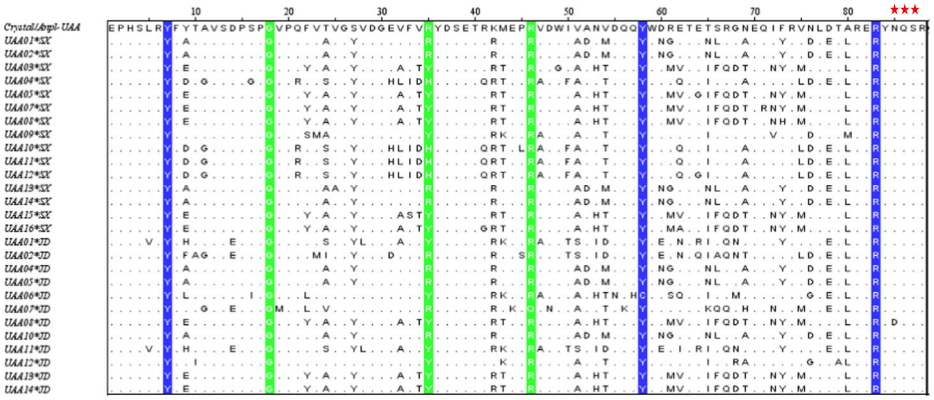
Fig. 1 Alignment of the 27 UAA sequences from Shaoxing (SX) and Jinding (JD) ducks obtained in the present study. The UAA allele codes for signal peptide, α1, α2, and α3 domains, followed by transmembrane and cytoplasmic domains. The conserved amino acids that form disulfide bonds and the N-linked glycosylation site are denoted by red numbers and by asterisks, respectively. The conserved amino acids involved in peptide binding are highlighted in blue. The residues interacting with β2m and CD8 molecules are highlighted in green and red, respectively

In our previous study, *Anpl*-MHC I genes were divided into 13 groups according to full-length amino acid homology > 91% (Zhang et al. 2017). When the UAA alleles of SX and JD ducks were introduced in the homology analysis, the number of groups rose to 17, including four novel groups. Although alleles expressed in the two duck breeds belonged to different groups (Table 2) and sequences isolated from individuals displayed highly variable polymorphism (Fig. 1), genes within the same group likely have similar peptide-binding properties given their high sequence homology. To further strengthen this hypothesis, the presence of HVSs, amino acid composition of epitope-binding pockets, and specificity of anchor residues were examined within each gene group. Because the Wu-Kabat index values within each group were lower than 4, there were no HVSs in their members; the amino acid compositions of the six pockets were similar within one group; through co-refolding of *Anpl*-MHC I molecules and peptides, the binding motif of peptides had individual rules in the same group (data not shown). Thus, to date, gene groups are a more appropriate classification for genetic structure and gene function studies than strain groups (Yan et al. 2005). Determining the distribution in duck lines of alleles with different peptide-binding characteristics is another purpose of the classification using allelic groups, which is helpful to develop resistant breeds to certain diseases and restricted CTL epitope vaccines (Zinkernagel et al. 1985; Koch et al. 2007; Nejentsev et al. 2007; Zhang et al. 2012).

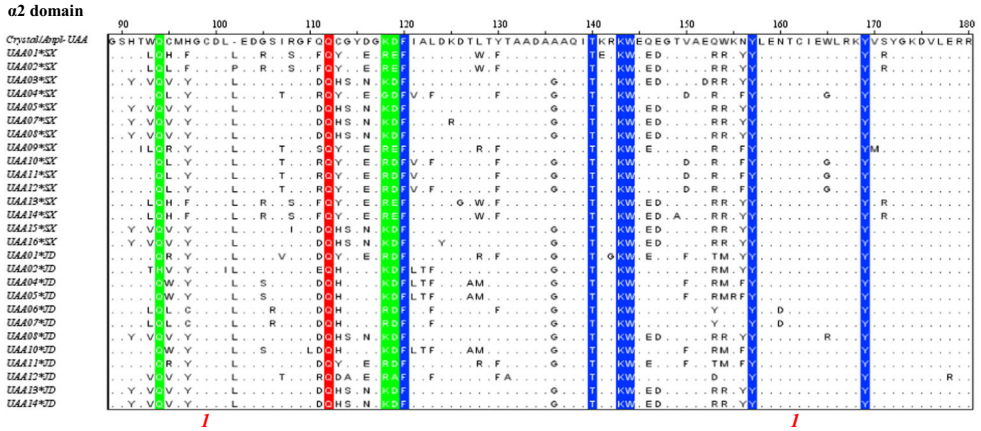
At both the breed and individual levels, allelic diversity was higher in JD than in SX ducks. This might be closely related to the habitat and feeding mode of these breeds. While SX ducks are fed in confined environments and the food source is unitary, JD ducks are stocked on the beach, with open breeding environment, and multiple food sources. Additionally, artificial selection for JD duck has been carried out repeatedly since 1958 (Xiao and Deng 2005), which might have increased selection pressure to generate high genetic diversity and polymorphism.

According to the crystal structure of *Anpl*-UAA*01 (Wu et al. 2017), UAA molecules bind peptides through six pockets in the PBG formed by the two α helixes and the β sheet in α1 and α2 domains, similar to that reported for human and chicken MHC I (Koch et al. 2007; Petersen et al. 2009). Each MHC molecule has individual procedures regarding the peptides presented in the PBG due to the high polymorphism of the MHC I sequences

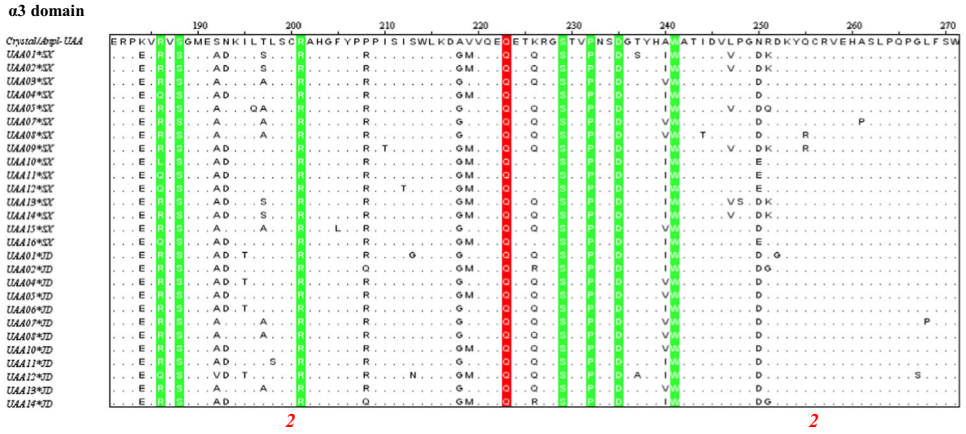
α1 domain



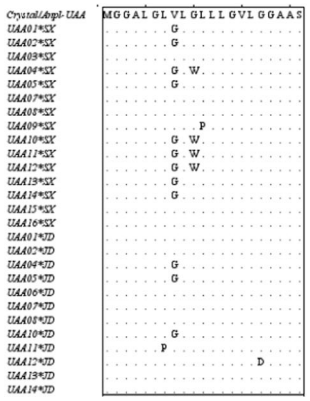
α2 domain



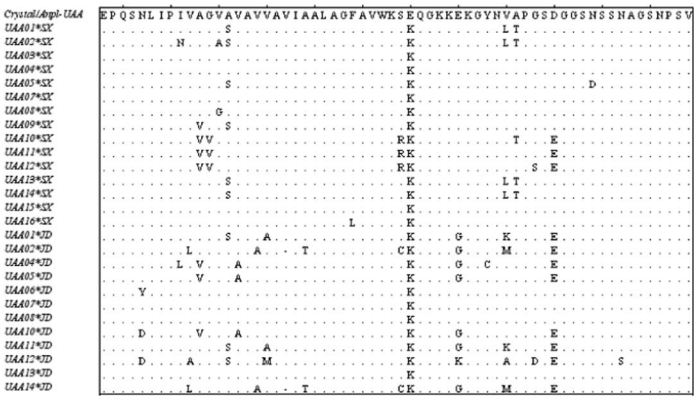
α3 domain



Signal peptide



TM /CY domain



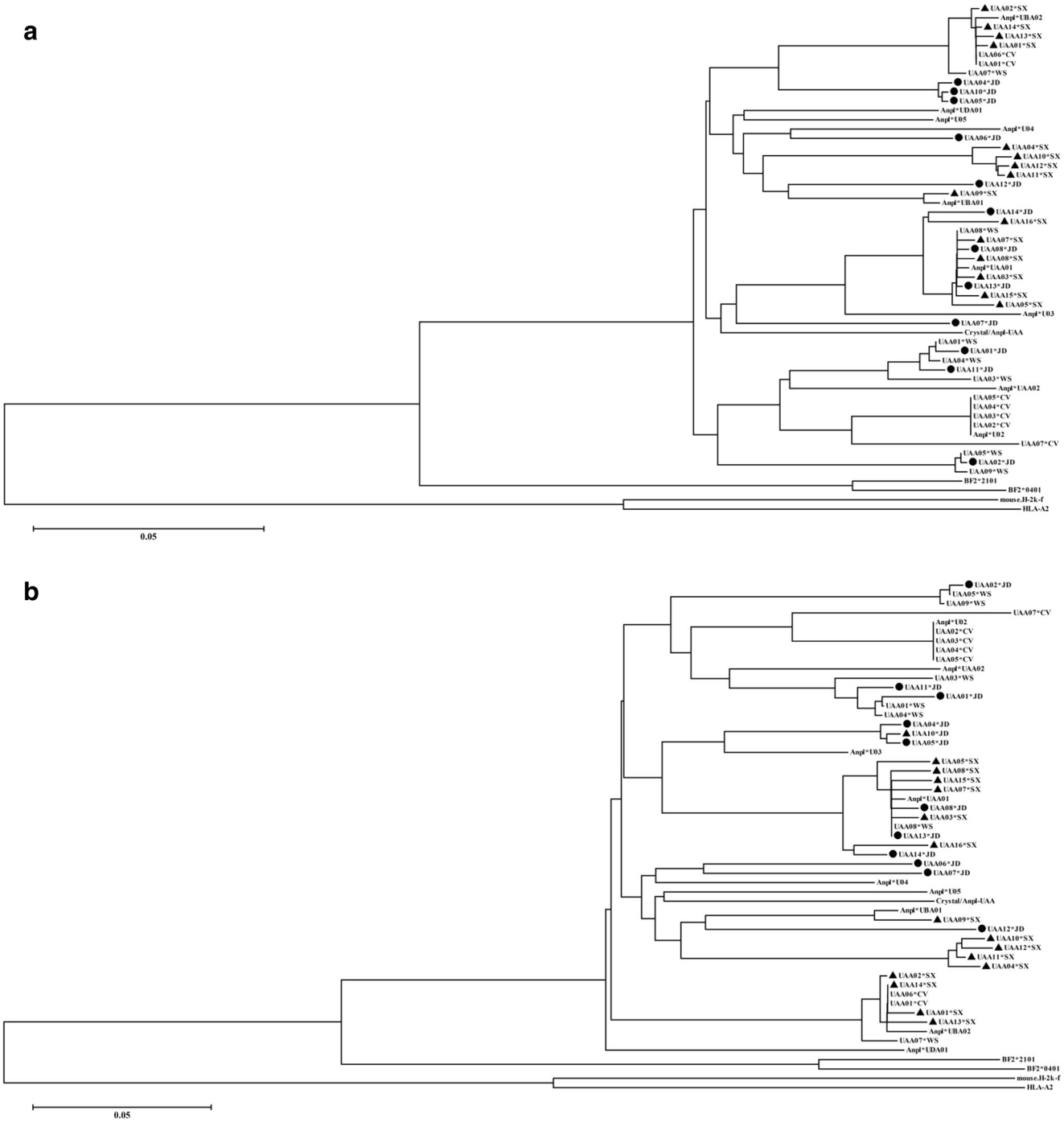


Fig. 2 Phylogenetic tree of *Anpl*-MHC I molecules constructed in MEGA6 using the neighbor-joining method and p-distance. Bootstrap values were obtained from 1000 replicates. The UAA sequences of Shaoxing (SX) and Jinding (JD) ducks identified in the present study are indicated by filled triangles and circles, respectively. **a** Phylogenetic tree constructed using the extracellular domain of all *Anpl-MHC I* nucleotide sequences obtained together with *Anpl-MHC I* sequences from NCBI. **b** Phylogenetic tree constructed with the amino acid sequences deduced from all obtained and retrieved *Anpl-MHC I* nucleotide sequences

(Falk et al. 1991). The HVSs of SX and JD ducks were mainly clustered at the α1 and α2 domains, especially in the four pockets (B, C, E, and F) that directly interact with peptides, and this benefits the diversity of peptide presentation. Moreover, *Anpl*-UAA*01 restricted peptides were selected relying on the B and F pockets, and in the present study, nearly half of the HVSs were located in these two pockets. Different *Anpl*-MHC I alleles had different peptide-binding specificities, and HVSs might influence the characteristics of peptide presentation. However, in the collective analysis of the 27 UAA allele

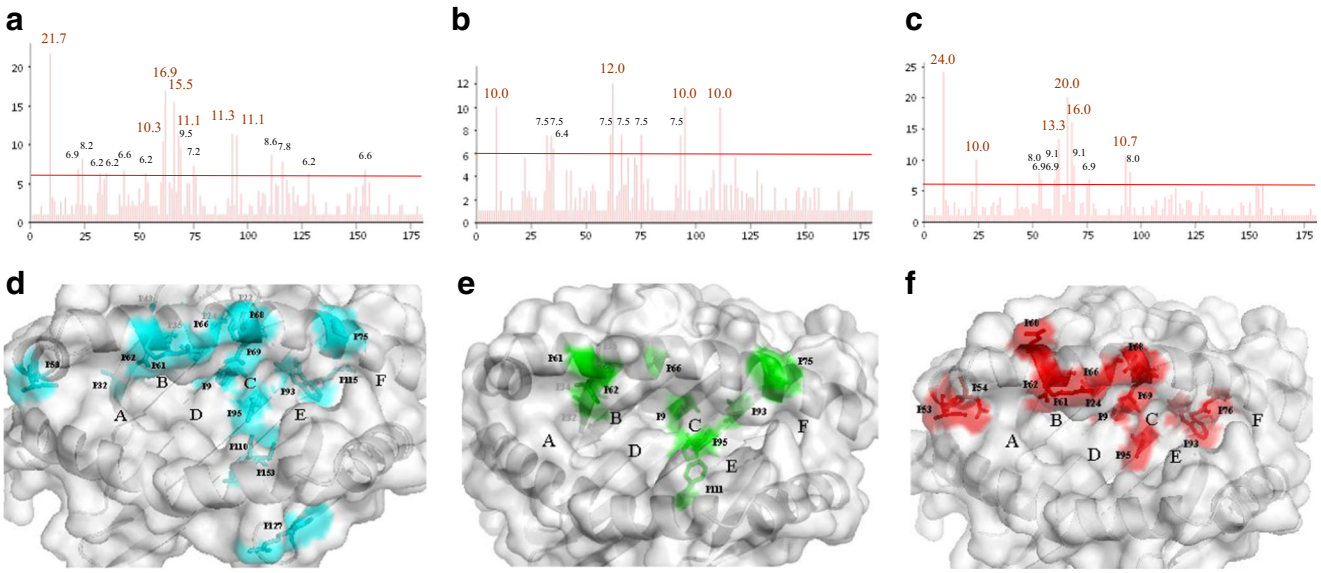


Fig. 3 Wu-Kabat plot of amino acid variability and highly variable sites located in the three-dimensional structures of the peptide binding groove (PBG) of the UAA. **a** Variability plots of the 27 UAA alleles. **b**, **c** Variability plots of the Shaoxing (SX) and Jinding (JD) duck breeds, respectively. **d** Cyan indicates the highly polymorphic sites in the structure of *Anpl*-UAA*01 (PDB code 5GJX). **e**, **f** Green and red indicate highly polymorphic sites from the SX and JD ducks in the structure of *Anpl*-UAA*01 (PDB code 5GJY). Numbers in **a**, **b**, and **c** indicate the variability scores of HVSs; combinations of letters and numbers in **d**, **e**, and **f** indicate the locations of HVSs

Table 3 Highly variable sites' variability scores within α1 and α2 domains of UAA in Shaoxing (SX) and Jinding (JD) ducks

Position	SX and JD	SX	JD
9	21.7	10	24
22	6.9		
24	8.2		10
32	6.2	7.5	
34		7.5	
35	6.2	6.4	
43	6.6		
53	6.2		8
54			6.9
60			6.9
61	10.3	7.5	9.1
62	16.9	12	13.3
66	15.5	7.5	20
68	11		16
69	9.5		9.1
75	7.2	7.5	
76			6.9
93	11.3	7.5	10.7
95	11.1	10	8
111	8.6	10	
116	7.8		
128	6.2		
154	6.6		

sequences, over 50% of the HVSs showed an offsetting effect with variability values lower than that found in each or in both breeds (Table 3). This demonstrated that alleles with similar sequences were commonly expressed in all duck lines in agreement with the results of allelic groups. Overall, the several UAA alleles and their high degree of polymorphism collectively contribute to a wide range of presented peptides, which is advantageous for pathogen resistance in ducks.

Conclusion

The polymorphism of UAA allele sequences analyzed in SX and JD ducks and their division according to groups evidenced their practical value for enriching *Anpl*-MHC I gene information and evaluating the distribution of UAA alleles in different Chinese duck breeds. These results might also accelerate effective diagnostics and vaccine development against infectious diseases.

Acknowledgements This study was carried out with financial support from the 13th 5-Year National Key Research Project of China (2016YFD0500106), the National Science and Technology Pillar Program during the 12th 5-Year Plan Period of China (2015BAD12B03), the Taishan Scholar Project Funds, Shandong Province (ts201511069), and the Agricultural Science and Technology Innovation Project Funds of Shandong Academy of Agricultural Science (CXGC2018E10 and CXGC2016B14).

Author's contributions Conceived and designed the experiments: L Zhang, D Lin, and J Wu. Suggested methodologies: L Zhang, D Lin, S Yu, W Jiang, and J Wu. Gene clone and analysis: S Yu, J Bai, and W Jiang. Allelic group division and HVS location: L Zhang and Y Huang. Provided the resources: D Lin, W Su, and S Yang. Wrote the paper: L Zhang and D Lin. Reviewed the paper: W Jiang and J Wu. All authors approved the final version of the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All animal experiments were performed in strict accordance with the guidelines of the Institutional Animal Care and Use Committee of Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences (IACC20060101, January 1, 2006).

References

Bacon LD, Witter RL (1993) Influence of B-haplotype on the relative efficacy of Marek's disease vaccines of different serotypes. Avian Dis 37(1):53–59

Bacon LD, Witter RL, Crittenden LB, Fadly A, Motta J (1981) B-haplotype influence on Marek's disease, Rous sarcoma, and lymphoid leucosis virus-induced tumors in chickens. Poult Sci 60(6):1132–1139

Bingham J, Green DJ, Lowther S, Klippel J, Burggraaf S, Anderson DE, Wibawa H, Hoa DM, Long NT, Vu PP, Middleton DJ, Daniels PW (2009) Infection studies with two highly pathogenic avian influenza strains (Vietnamese and Indonesian) in Pekin ducks (*Anas platyrhynchos*), with particular reference to clinical disease, tissue tropism and viral shedding. Avian Pathol 38(4):267–278

Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. Annu Rev Biochem 59(1):253–288

Borbulevych OY, Piepenbrink KH, Baker BM (2011) Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. J Immunol 186(5):2950–2958

Chan WF, Parks-Dely JA, Magor BG, Magor KE (2016) The minor MHC class I gene UDA of ducks is regulated by let-7 microRNA. J Immunol 197(4):1212–1220

Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351(6324):290–296

Fan SH, Wang YL, Wang S, Wang X, Wu YN, Li ZB, Zhang NZ, Xia C (2018) Polymorphism and peptide-binding specificities of porcine major histocompatibility complex (MHC) class I molecules. Mol Immunol 93:236–245

Fleming-Canepa X, Jensen SM, Mesa CM, Diaz-Satizabal L, Roth AJ, Parks-Dely JA, Moon DA, Wong JP, Evseev D, Gossen DA, Tetrault DG, Magor KE (2016) Extensive allelic diversity of MHC class I in wild mallard ducks. J Immunol 197(3):783–794

Goto RM, Wang YJ, Taylor RL, Wakenell PS, Hosomichi K, ShiinaT BCS, Briles WE, Miller MM (2009) BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. Proc Natl Acad Sci USA 106(39):16740–16745

Hazes B, Dijkstra BW (1988) Model building of disulfide bonds in proteins with known three-dimensional structure. Protein Eng 2(2): 119–125

Hepkema BG, Blankert JJ, Albers GA, Tilanus MG, Egberts E, van der Zijpp AJ, Hensen EJ (1993) Mapping of susceptibility to Marek's disease within the major histocompatibility (B) complex by refined typing of white leghorn chickens. Anim Genet 24(4):283–287

Jondal M, Schirmbeck R, Reimann J (1996) MHC class I-restricted CTL responses to exogenous antigens. Immunity 5(4):295–302

Kabat EA, Wu TT, Bilofsky H (1997) Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. J Biol Chem 252(19): 6609–6616

Kane K, Clark WR (1984) The role of class I MHC products in polyclonal activation of CTL function. J Immunol 133(6):2857–2863

Kaufman J (1999) Co-evolving genes in MHC haplotypes: the "rule" for non-mammalian vertebrates? Immunogenetics 50(3–4):228–236

Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjødt K (1992) Different features of the MHC class I heterodimer have evolved at different rates: chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. J Immunol 148(5):1532–1546

Kaufman J, Volk H, Wallny HJ (1995) A "minimal essential Mhc" and an"unrecognized Mhc": two extremes in selection for polymorphism. Immunol Rev 143(1):63–88

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401(6756):923–925

Klein J, Rammensee HG, Nagy ZA (1983) The major histocompatibility complex and self and non-self differentiation through the immune system. Die Naturwissenschaften 70(6):265–271

Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny HJ, van Hateren A, Hunt L, Jacob JP, Johnston F, Marston DA, Shaw I, Dunbar PR, Cerundolo V, Jones EY, Kaufman J (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity 27(6):885–899

Li N, Wang Y, Li R, Liu JY, Zhang JZ, Cai YM, Liu SD, Chai TJ, Wei LM (2015) Immune responses of ducks infected with duck Tembusu virus. Front Microbiol 6:425

Liang QL, Luo J, Zhou K, Dong JX, He HX (2011) Immune-related gene expression in response to H5N1 avian influenza virus infection in chicken and duck embryonic fibroblasts. Mol Immunol 48(6–7): 924–930

Lie WR, Myers NB, Connolly JM, Gorka J, Lee DR, Hansen TH (1991) The specific binding of peptide ligand to Ld class I major histocompatibility complex molecules determines their antigenic structure. J Exp Med 173(2):449–459

Liu WJ, Kong ZJ, Li HY, Liu ZQ, Yang SH, Huang YY, Zhang XM, Zhang L, Wu JQ (2017) Comparison of host innate immune response of chicken (*Gallus gallus*) and duck (*Anas platyrhynchos*) infected with duck-origin Newcastle disease virus. J Agric Biotechnol 25(2):307–315

McKenzie LM, Pecon-Slattery J, Carrington M, O'Brien SJ (1999) Taxonomic hierarchy of HLA class I allele sequences. Genes Immun 1(2):120–129

McMichael AJ, Gotch FM, Noble GR, Beare PA (1983) Cytotoxic T-cell immunity to influenza. N Engl J Med 309(1):13–17

Mesa CM, Thulien KJ, Moon DA, Veniamin SM, Magor KE (2004) The dominant MHC class I gene is adjacent to the polymorphic TAP2 gene in the duck, Anas platyrhynchos. Immunogenetics 56(3):192–203

Moon DA, Veniamin SM, Parks-Dely JA, Magor KE (2005) The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed MHC genes. J Immunol 175(10):6702–6712

Nejentsev S, Howson JMM, Walker NM, Szeszko J, Field SF, Stevens HE, Reynolds P, Hardy M, King E, Masters J, HulmeJ MLM, Smyth D, Bailey R, Cooper JD, Ribas G, Campbell RD, Clayton DG, Todd JA, Welcome Trust Case Control Consortium (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature 450(7171):887–892

Petersen J, Wurzbacher SJ, Williamson NA, Ramarathinam SH, Reid HH, Nair AKN, Zhao AY, Nastovska R, Rudge G, Rossjohn J, Purcell AW (2009) Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. Proc Natl Acad Sci U S A 106(8):2776–2781

Plachy J, Benda V (1981) Location of the gene responsible for Rous sarcoma regression in the B-F region of the B complex (MHC) of the chicken. Folia Biol (Praha) 27(5):363–368

Plachy J, Pink JR, Hala K (1992) Biology of the chicken MHC (B complex). Crit Rev Immunol 12(1–2):47–79

Rawle FC, Knowles BB, Ricciardi RP, Brahmacheri V, Duerksen-Hughes P, Wold WS, Gooding LR (1991) Specificity of the mouse cytotoxic T lymphocyte response to adenovirus 5. E1A is immunodominant in H-2b, but not in H-2d or H-2k mice. J Immunol 146(11):3977–3984

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, Wiles MV, Milne S, Beck S, Kaufman J (2007) Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol 178(9):5744–5752

Takeshima SN, Sarai Y, Saitou N, Aida Y (2009) MHC class II DR classification based onantigen-binding groove natural selection. Biochem Biophys Res Commun 385(2):137–142

Walker BA, van Hateren A, Milne S, Beck S, Kaufman J (2005) Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. Immunogenetics 57(3–4):232–247

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjodt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) 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 103(5):1434–1439

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191

Wu YN, Wang JY, Fan SHH, Chen R, Liu YJ, Zhang JH, Yuan HY, Liang RY, Zhang NZ, Xia C (2017) Structural definition of duck major histocompatibility complex class I molecules that might explain efficient cytotoxic T lymphocyte immunity to influenza a virus. J Virol 91(14):e02511–e02516

Xia C, Lin CY, Xu GX, Hu TJ, Yang TY (2004) cDNA cloning and genomic structure of the duck (*Anas platyrhynchos*) MHC class I gene. Immunogenetics 56(4):304–309

Xiao GM, Deng YB (2005) Duck cultivation. Hunan science and technology press, Hunan

Yan RQ, Li XS, Yang TY, Xia C (2005) Characterization of BF2 and β2m in three Chinese chicken lines. Vet Immunol Immunopathol 108(3):417–425

Zhang W, Collins A, Morton NE (2004) Does haplotype diversity predict power for association mapping of disease susceptibility? Hum Genet 115(2):157–164

Zhang JH, Chen Y, Qi JX, Gao F, Liu Y, Liu J, Zhou X, Kaufman J, Xia C, Gao GF (2012) Narrow groove and restricted anchors of MHC class I molecule BF2 *0401 plus peptide transporter restriction can explain disease susceptibility of B4 chickens. J Immunol 189(9):4478–4487

Zhang L, Liu WJ, Wu JQ, Xu ML, Kong ZHJ, Huang YY, Yang SHH (2017) Characterization of duck (*Anas platyrhynchos*) MHC class **I** gene in two duck lines. J Genet 96(2):371–375

Zhou M, Xu Y, Lou Z, Cole DK, Li X, Liu Y, Tien P, Rao Z, Gao GF (2004) Complex assembly: crystallization and preliminary X-ray crystallographic studies of MHC H-2Kd complexed with an HBV-core nonapeptide. Acta Crystallogr F Struct Biol Commun 60(Pt8):1473–1475

Zinkernagel RM, Pfau CJ, Hengartner H, Althage A (1985) Susceptibility to murine lymphocytic choriomeningitis maps to class I MHC genes? A model for MHC/disease associations. Nature 316(6031):814–817