ORIGINAL PAPER

Characterization and locus-specific typing of MHC class I genes in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci

Alison Cloutier · James A. Mills · Allan J. Baker

Abstract A major challenge facing studies of major histocompatibility complex (MHC) evolution in birds is the difficulty in genotyping alleles at individual loci, and the consequent inability to investigate sequence variation and selection pressures for each gene. In this study, four MHC class I loci were isolated from the red-billed gull (*Larus scopulinus*), representing both the first characterized MHCI genes within Charadriiformes (shorebirds, gulls, and allies) and the first full-length MHCI sequences described outside Galloanserae (gamebirds + waterfowl). Complete multilocus genotypes were obtained for 470 individuals using a combination of reference-strand conformation analysis and direct sequencing of gene-specific amplification products, and variation of peptide-binding region (PBR) exons was surveyed for all loci. Each gene is transcribed and has conserved sequence features characteristic of antigen-presenting MHCI molecules. However, higher allelic variation, a more even allele frequency distribution, and evidence of positive selection acting on a larger number of PBR residues suggest that only one locus (Lasc-UAA) functions as a major classical MHCI gene. Lasc-UBA, with more limited variation and PBR motifs that encompass a subset of Lasc-UAA diversity, was assigned a putative minor classical function, whereas the divergent and largely invariant binding-groove motifs of Lasc-UCA and -UDA are suggestive of nonclassical loci with specialized ligand-binding roles.

A. Cloutier (✉) · A. J. Baker
Department of Ecology and Evolutionary Biology,
University of Toronto,
Toronto, ON, Canada M5S 3B2
e-mail: alison.cloutier@utoronto.ca

J. A. Mills
10527 A Skyline Drive,
Corning, NY 14830, USA

A. J. Baker
Department of Natural History, Royal Ontario Museum,
100 Queen's Park,
Toronto, ON, Canada M5S 2C6

Introduction

The major histocompatibility complex (MHC) encompasses families of highly polymorphic genes encoding antigen-presenting molecules with a central role in the acquired immune response. MHC class I genes, in association with β_2-microglobulin, form molecules that bind peptides derived from intracellular pathogens for display at the cell surface where recognition by CD8+ T cells initiates an immune response (Klein 1986). Residues of the peptide-binding region (PBR) that is formed from the $\alpha1$ and $\alpha2$ domains encoded by the second and third exons, respectively, of MHCI genes determine the specificity of peptide binding. Classical MHC class Ia genes are widely expressed and are typified by high variability and positive selection acting upon substitutions at PBR sites, in addition to a large number of alleles maintained by the action of balancing selection. In contrast, nonclassical MHC Ib genes usually show reduced expression or limited tissue distribution, are less polymorphic than class Ia genes, and may

have restricted or modified functions (Hughes and Nei 1988, 1989; Rodgers and Cook 2005).

Avian MHCI gene function and evolution is most fully described for the chicken (*Gallus gallus*, order Galliformes), where the BF–BL region of the "minimal essential" MHC (the B locus of microchromosome 16) contains two classical MHCI genes—the major, or predominantly expressed BF2 locus, and the less variable minor gene, BF1 (Kaufman 1999; Kaufman et al. 1999a, b). Additional MHCI gene family members, including at least one functional nonclassical locus, localize to the secondary MHC-Y region that segregates independently of the classical MHC (Miller et al. 1994; Afanassieff et al. 2001; Hunt et al. 2006; Hee et al. 2010). Expanded MHCI gene numbers are described for two other galloanseriforms: the Japanese quail (*Coturnix japonica*, order Galliformes) and the mallard duck (*Anas platyrhynchos*, order Anseriformes). However, only one (duck) or two (quail) genes exhibit the hallmarks of classical MHCI loci (with predominant expression of a single quail locus, Coja-E; Shiina et al. 2006), which suggests that a functional minimization of this gene family, at least in terms of classical gene function, could be common to birds (Shiina et al. 2004; Moon et al. 2005). Predominant expression of a major MHCI locus in galloanseriforms is attributed to tight linkage with the transporter associated with antigen processing genes TAP1 and 2, which is believed to result in the coevolution of alleles with coordinated functional specificities within haplotypes (Kaufman 1999; Kaufman et al. 1999a; Mesa et al. 2004). Expanded numbers of MHCI genes are also found in passerines (perching birds, order Passeriformes), but functional categorization of all gene family members is not currently available, and it is not yet known if a major MHC class I locus is typical within this group (Westerdahl 2007).

Despite continuing investigation of avian MHCI genes, and an increasing interest in their use as markers of adaptive genetic variation in evolutionary ecology research (Edwards et al. 2000; Piertney and Oliver 2006; Babik 2010), there remain many important gaps in our knowledge of MHCI multigene family evolution in birds. Most studies have focused upon the isolation of cDNA transcripts or the variation within the polymorphic PBR exons, with the result that no full-length sequences have been described for non-galloanseriforms. Limited DNA sequence information, combined with differences in gene copy number, the presence of pseudogenes, and the inability to identify orthologous loci from sequence similarity in any but the most closely related species (Hughes and Nei 1989; Shawar et al. 1994) necessitates de novo characterization of MHCI genes in each studied species. However, locus-specific amplification in birds is commonly hampered by the existence of highly con-

served sequences surrounding the PBR, and locus assignment based upon PBR sequence similarity is often impeded by conflicting evolutionary signals due to recent duplication, recombination and/or concerted evolution (Edwards et al. 2000; Hess and Edwards 2002; Westerdahl 2007; Babik 2010). Consequently, the evolutionary forces acting upon each gene cannot be disentangled from a larger picture of polymorphism and selective pressures averaged across all loci, and the relative contributions of classical versus nonclassical genes to overall variation cannot be evaluated. Finally, current knowledge of avian MHCI genes remains taxonomically restricted, which limits the potential of comparative methods and a broader understanding of the evolutionary processes acting upon these genes.

In this study, we aimed to isolate full genomic and cDNA MHCI sequences from the red-billed gull (*Larus scopulinus*, order Charadriiformes), and to survey the variation in PBR-encoding exons for all loci. The long-term study of red-billed gull breeding ecology (Mills 1989) makes this species an excellent choice as the pioneer for MHCI research in charadriiforms (shorebirds, gulls, and allies), as a large number of known family groups are available to facilitate the assignment of alleles to loci through segregation analyses. Characterization of the variation, selection, and expression of individual loci was pursued to assess whether a major classical MHCI locus exists in the red-billed gull. Ancillary methodological goals were to investigate the usefulness of noncoding sequence variation for the design of locus-specific amplification strategies, and to evaluate the applicability of reference-strand conformation analysis (RSCA, supplemental Fig. S1) as a tool for screening MHCI variation in a nonmodel bird species with no existing genetic resources. RSCA is emerging as a popular option for MHC genotyping (e.g., Noakes et al. 2003; Pratt et al. 2006; Worley et al. 2008; Lenz et al. 2009; Babik 2010). However, its use in avian MHC studies is largely limited to the red jungle fowl (*G. gallus gallus*; Worley et al. 2008; Gillingham et al. 2009), where extensive research of its domesticated descendant, the chicken, provides a priori knowledge of MHC organization and gene content that is unavailable for other species.

Materials and methods

Standard PCR and sequencing conditions

PCR reactions contained final concentrations of 1X commercially supplied buffer (20 mM Tris–HCl pH 8.4, 50 mM KCl), 1.5 mM $MgCl_2$, 0.2 mM each deoxyribonucleotide triphosphate, 0.5 µM each of forward and reverse primers, 5% dimethyl sulfoxide, 1/3 U Platinum Taq DNA polymerase (Invitrogen), and 50–100 ng template DNA in a

25-μL final reaction volume. PCR cycling conditions were: initial denaturation at 94°C for 5 min followed by 35 cycles of denaturation at 94°C for 45 s, annealing for 45 s, and extension at 72°C, and a final elongation step at 72°C for 5 min. A minimum extension time of 30 s was used; for longer PCR fragments, extension times of 1 min per kilobase amplified were used.

Following electrophoresis, PCR products were recovered from agarose gels by band excision and filtered pipette tip centrifugation (Dean and Greenwald 1995). Standard sequencing reactions used 4 μL of recovered DNA with Big Dye Terminator v. 3.1 chemistry (Applied Biosystems), and fragments were resolved by automated capillary electrophoresis on an ABI 3100.

Isolation of MHCI cDNA transcripts

A directionally cloned, pooled cDNA library was commercially prepared (Bio S&T, Montreal, Canada) from spleen tissue harvested from a single adult female, with each library pool consisting of 1,000 clones carrying primary transcripts ligated into pBlueScript SK⁻ vector. Hybridization-based screening used a 167-bp probe amplified from a conserved region of exon 4. Following PCR amplification with primers MHC1AF/MHC1AR (Table 1), probe DNA was biotin-labelled using the NEBlot Phototope Kit (New England BioLabs). Hybridizations of colony-growth membranes were performed at 65°C, with subsequent detection of cDNA pools producing positive signals to exon 4 probe using the Phototope-Star Detection Kit for Nucleic Acids (New England BioLabs). DNA from each positive cDNA pool was amplified with primers MHC1OF/MHC1KR and sequenced to identify the MHCI allele present in the pool. Full transcript sequences were obtained from five cDNA pools per allele using primer combinations MHC1OF/M13F and MHC1SR/M13R.

PCR-based cDNA library screening was used to screen for alleles isolated from genomic DNA that were not detected during hybridization-based cDNA library screening. Amplification with gene-specific primer combinations A3F/R (Lasc-UBA), A4F/R (Lasc-UCA), and A5-6F/R (Lasc-UDA; Table 1) was used to identify cDNA pools uniquely containing transcripts from each of these loci. Full transcript sequences were subsequently isolated from three such pools per gene using each gene-specific primer in combination with either M13F or M13R and an annealing temperature of 60°C in all instances.

Isolation of genomic MHCI sequences

A ~1,500-bp fragment spanning exons 2–5 and a 1,844-bp fragment covering exon 3 to the 3′UTR were amplified using primer combinations MHC1OF/MHC1KR and MHC1LF/MHC1ER, respectively, from the same individual used for cDNA library construction. PCR products were cloned into pCR 2.1 vector using The Original TA Cloning Kit (Invitrogen), and inserts amplified with M13F/M13R primers and sequenced according to standard methods. Sequences extending to the 5′ and 3′ ends of cloned fragments were obtained from shotgun sequencing and primer walking in cosmid clones containing each MHCI gene (A. Cloutier, unpublished data) from the same individual used for the cDNA and genomic cloning studies.

Screening of peptide-binding region variation

Sample collection and DNA extraction

Blood samples were collected from individually colour- and metal ring-banded gulls at a breeding colony in Kaikoura, New Zealand during the 1992–2004 breeding seasons. DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). MHCI variation was screened in 470 individuals, consisting of 45 family groups where chicks had known maternity and paternity and 71 family groups with known maternity only (32 female–female social pairs; 39 male–female pairs with broods of mixed paternity or all extrapair young), with 1.75 ± 0.82 (mean ± SD) genotyped offspring per family group. Parentage assignments made from field observations were verified using multilocus microsatellite genotyping (J. Mills and A. Cloutier, unpublished data).

Reference-strand conformation analysis

A ~1,300-bp fragment spanning the 3′ end of intron 1 through the 5′ end of intron 3 was simultaneously amplified from all MHCI genes with primers MHCI-F5/EX3R1 in 15 individuals. Following TA cloning and M13F/R amplification of insert DNA, 8–12 clones/individual were sequenced.

Five alleles (Lasc-UAA*03, UAA*04, UBA*01, UCA*01, and UDA*07) were tested as potential fluorescently labelled references (FLRs). A 350-bp fragment spanning exon 3 was amplified from cloned DNA with primers EX3F1/EX3R2 using the standard conditions outlined above, except: (1) the forward primer was 5′ end labelled with VIC dye (Applied Biosystems), and (2) a 5:1 ratio of forward to reverse primer was used. FLR products were purified using the QIAquick PCR Purification Kit (Qiagen) to remove unincorporated primers and stored at 4°C until use.

Products for simultaneous screening of MHCI genes were amplified from genomic DNA using primers EX3F1/EX3R2. Additionally, gene-specific RSCA was performed for Lasc-UCA and Lasc-UDA to allow preliminary locus assignment (i.e., assignment of alleles to Lasc-UCA, -UDA, or -UAA and -UBA). Initial PCRs used standard conditions

Table 1 Primers used in MHC class I gene characterization and survey of peptide-binding region variation

Procedure	Primers	Sequence (5′→3′)	Region	Size (bp)	Tm (°C)
cDNA					
Hybridization probe	MHC1AF	GCTCACGGCTTCTACCCG	Exon 4	167	63.5
	MHC1AR	CGGCACTGGTACTTGTCC			
Identification of cDNA allele	MHC1OF	GVRSCGCTACAACCAGAGC	Exon 2–5	601	60
	MHC1KR	CACGATGGSSMNCAGGTTGGACT			
Isolation of cDNA transcript	MHC1OF	GVRSCGCTACAACCAGAGC	Exon 2–3′UTR	1,608	59
	M13F	CACGACGTTGTAAAAACGAC			
	MHC1SR	AGGTTGGCCGCCATCCAGTC	Exon 2–5′ end	Variable[a]	60
	M13R	CAGGAAACAGCTATGAC			
cDNA PCR screening	A3F	CAATCAGCAGGTTGACCGCGTCAG	Exon 2–3	149	65
	A3R	TCCCGTCGTAGCGTACTGCCGAA			
	A4F	TGTGAATCTGGAGATACTGGGGGT	Exon 2–3	264	65
	A4R	GCAGGTGTTCTTCACATAGTGCTGCTT			
	A5-6 F	ACTGGATGGCGGCCAACCTGGACCA	Exon 2–3	338	63
	A5-6R	AATCTCCAGACAGTGCTTGC			
Genomic DNA					
TA cloning	MHC1OF	GVRSCGCTACAACCAGAGC	Exon 2–5	1,417–1,514[a]	60
	MHC1KR	CACGATGGSSMNCAGGTTGGACT			
	MHC1LF	AGTGGCTGRGGAAATACGTGAGCT	Exon 3–3′UTR	1,844	60
	MHC1ER	TATCGAACATGACGTTCACG			
RSCA					
Reference allele cloning	MHCI-F5	ACCCCCCGTGTCCCCGCAGG	Intron 1–3	1,278–1,375[a]	68
	EX3R1	CGGTGAGGGACGGCACACGCACCCG			
Exon 3 screening	EX3F1	AGCCCGGCTCCTGGTCGCGTTCCAG	Intron 2–3	350	68.6
	EX3R2	ACACGCACCCGCGCTGGCCGCTCCCCAG			
Exon 3 gene-specific PCR	Lin145F	AGATGAGACCCCCAGAAGCAGAGGGGCTCCCC	Intron 2–3′UTR	2,171	TD 70→66[b]
	Lin45-3UTR2	CGTTTATTGAGCGTGCCAGGAATCGTGC			
	Lin145F	AGATGAGACCCCCAGAAGCAGAGGGGCTCCCC	Intron 2–3′UTR	2,178	
	Lin1-3UTR1	GCGTTCCAAGGATTTCCAATGTGATGCA			
	Lin2F	GTCGGCTCCTGTCCACCCCAGCCAGTGCAT	Intron 2–3	849	68.6
	EX3R1	CGGTGAGGGACGGCACACGCACCCG			
	Lin3F	ATGCCGTGGGACGGGAGAGTATCCAAGCCC	Intron 2–3	974	68.6
	EX3R1	CGGTGAGGGACGGCACACGCACCCG			
Exon 2 gene-specific PCR	MHCI-F5[c]	ACCCCCCGTGTCCCCGCAGG	Intron 1		
	Lin145R	GGGGAGCCCCTCTGCTTCTGGGGGTCTCATC	Intron 2	658	68.6
	Lin45BOV	AGGTACTGCTTCCACCGCTCGGCTTCAGC	Exon 3	1,245	TD 70→62[b]
	Lin45FX	CAGGTGTTCTGCAGGTACTGCTTCAACT	Exon 3	1,257	
	Lin1-EX3R	CAGCCATACATGGTCTGCACTGTGTGAA	Exon 3	1,066	65
	Lin2R	ATGCACTGGCTGGGGTGGACAG	Intron 2	543	68.6
	Lin3R	GGCATCCCGTCCCAGCCCGCGCTCACT	Intron 2	309	68.6

TD touchdown PCR profile

[a] Size differences result from variable extension during cDNA synthesis or variable intron sizes among loci

[b] Touchdown PCR profile, (−1°C/cycle) followed by 30 cycles at final touchdown annealing temperature

[c] Forward primer used in all exon 2 amplifications

with primers Lin2F/EX3R1 (Lasc-UCA) or Lin3F/EX3R1 (Lasc-UDA). PCR products were diluted 100-fold, and 2.5 μL of diluted DNA was used in a reamplification with primers EX3F1/EX3R2 and a touchdown profile of 71→68.5°C (−0.5°C/cycle), followed by 30 cycles at 68.5°C annealing temperature.

Two microliters of a 1/25 dilution of FLR and 2.5 μL of genomic PCR product were mixed, and hybridized as follows: 95°C for 10 min, decreased to 55°C (−1°C/sec),

55°C for 15 min, decreased to 4°C at maximum ramp speed, and 4°C for 15 min. Following hybridization, 15 μL of H_2O was added to each reaction. Sample plates were prepared for processing on an ABI 3100 by combining the following in each well: 3 μL diluted hybridization mix, 0.3 μL LIZ500 size standard (Applied Biosystems), and 11.7 μL H_2O. Samples were run under nondenaturing conditions using 5% Genescan polymer (Applied Biosystems), 1X TBE buffer, and the following run module

settings: 30°C oven temperature, 1 kV injection voltage, 30 s injection time, 15 kV run voltage, and 2,700 s run time.

PBR exon sequencing

Direct sequencing of gene-specific amplification products was used to obtain sequences for all alleles identified during RSCA screening. PCRs spanning exon 3 used primer combinations: Lin145F/Lin45-3UTR2 (Lasc-UAA), Lin145F/Lin1-3UTR1 (Lasc-UBA), Lin2F/EX3R1 (Lasc-UCA), and Lin3F/EX3R1 (Lasc-UDA). Whenever possible, homozygous individuals (as determined from RSCA profiles) were used. Otherwise, different heterozygous combinations of alleles were amplified and sequenced, and haplotype assignments made using PHASE v. 2.1 software (Stephens and Scheet 2005).

Exon 2 sequences were also obtained for at least three copies of each RSCA allele. Lasc-UCA and -UDA were amplified with gene-specific primer combinations MHCI-F5/Lin2R and MHCI-F5/Lin3R, respectively. Lasc-UAA and -UBA could not be amplified independently for exon 2. Instead, homozygotes for each gene were co-amplified using primers MHCI-F5/Lin145R and haplotypes inferred using PHASE. Low-frequency alleles that could not be typed in this manner were amplified using allele-specific primer combinations MHCI-F5/Lin45BOV and MHCI-F5/Lin45FX (Lasc-UAA), and MHCI-F5/Lin1-EX3R (Lasc-UBA).

Genotyping of multilocus gene family members is prone to the formation of chimeric sequence artefacts during PCR and cloning procedures (Longeri et al. 2002; Lenz and Becker 2008). We therefore validated all described alleles by ensuring that cloned allele sequences coincided with alleles found during RSCA screening of the same individual, by sequencing at least three copies of each allele originating from different individuals using gene-specific PCR reactions, and by confirming all allele identities and locus assignments with segregation analyses in known family groups.

Data analysis

Analysis of sequence variation

ChromasPro v. 1.5 (Technelysium Pty Ltd.) was used for sequence editing and contig assembly. Sequences were aligned using Clustal W (Thompson et al. 1994) as implemented in BioEdit Sequence Alignment Editor v. 7.0.5.3 (Hall 1999). Sequence variability statistics were calculated in Molecular Evolutionary Genetics Analysis v. 4 (Tamura et al. 2007), including average rates of synonymous (d_S) and nonsynonymous (d_N) substitutions per site and codon-based Z tests of positive selection according to the Nei–Gojobori method with Jukes–Cantor correction for multiple substitutions (Nei and Gojobori 1986). Allele frequencies of unrelated adults (N=230) were calculated in Cervus v. 3.0.3 (Kalinowski et al. 2007), and Ewens–Watterson neutrality tests of allele frequency distributions (Ewens 1972; Watterson 1978) were conducted using Arlequin v. 3.5.1.2 (Excoffier et al. 2005).

Genetic Algorithm Recombination Detection (GARD, Kosakovski Pond et al. 2006) was used to identify recombination breakpoints. LDhat v. 2.1 (McVean and Auton 2007) was used to calculate the recombination parameter $4N_er$, population-scaled recombination rate (ρ), minimum number of recombination events (Rm), and the Watterson estimate of the mean population mutation rate (θ). Geneconv v. 1.81 (Sawyer 1999) was used to identify homogenized sequence tracts consistent with interlocus gene conversion.

Phylogenetic analysis

The use of a bifurcating treelike structure to describe sequence evolution of PBR-encoding exons may be unrealistic in the presence of conflicting phylogenetic signals resulting from gene conversion and/or recombination reported for avian MHCI genes (Alcaide et al. 2009; Promerová et al. 2009). Instead, phylogenetic network methods that allow the modeling of incompatibilities in the dataset as parallel splits may be warranted. SplitsTree v. 4.10 (Huson and Bryant 2006) was used to construct a neighbor-net phylogenetic network from concatenated exon 2 and 3 nucleotide sequences (540 bp total), using a GTR+Γ substitution model with empirical base frequencies and gamma distribution shape parameter α=0.17032, as recommended by FindModel (Tao et al. 2009), and 1,000 bootstrap resamplings of the dataset.

Inference of positive selection

OmegaMap v. 0.5 (Wilson and McVean 2006) implements a Bayesian strategy using reversible-jump Markov chain Monte Carlo (MCMC) simulations to co-estimate spatial variation in the selection parameter ($\omega=d_N/d_S$) and recombination rate (ρ). Analyses were run on concatenated exon 2 and 3 nucleotide datasets based on adult allele frequencies for each gene. Three independent MCMC chains of 500,000 iterations were run for each dataset, with ten random orderings of the data, a thinning interval of 100, and the first 50,000 iterations discarded as burn-in after graphical inspection of the raw output. Objective priors were set for the transition to transversion rate ratio (κ), synonymous transversion rate (μ), and rate of insertion to deletion (ϕ) by specifying an improper inverse distribution

and starting values of 3.0, 0.1, and 0.1, respectively. The spatial variation in selection (ω) and rate of recombination (ρ) were allowed to vary across sites, using an inverse distribution and starting values ranging from 0.01 to 100 for ω and 0.001–100 for ρ, and block sizes of 2 for both parameters.

Results

Isolation and characterization of red-billed gull MHCI genes

Four MHCI genes were isolated from genomic DNA of a single individual and named *MhcLasc*-UAA, *MhcLasc*-UBA, *MhcLasc*-UCA, and *MhcLasc*-UDA in accordance with proposed nomenclature (Klein et al. 1990), and with alleles at each locus numbered in order of their validation during the RSCA screening procedure. Full-length sequences were obtained for alleles Lasc-UAA*11, Lasc-UBA*01, and Lasc-UCA*01 (GenBank accessions HM008713–HM008715), and a partial sequence covering exon 2 through the end of the 3′UTR was obtained for allele Lasc-UDA*03 (HM008716).

Hybridization-based cDNA library screening detected 77 (of 1440) library pools producing strong positive signals. Allele Lasc-UAA*08 (HM015819) or Lasc-UAA*11 (HM015818) was identified in each positive pool, and full-length cDNA sequences were obtained for each allele. Full-length transcripts for alleles Lasc-UBA*01 (HM015820) and Lasc-UCA*01 (HM015821), and a partial transcript spanning exon 2 through to the poly-A tail for allele Lasc-UDA*03 (HM015822) were isolated by PCR-based re-screening of the library. In all instances, exon and UTR sequences obtained for cDNA transcripts were identical to consensus sequences derived from cloning and primer walking in genomic DNA.

The red-billed gull MHCI genes show the expected eight exons encoding the signal peptide, α1, α2, α3, transmembrane, and cytoplasmic domains. No frameshifts or premature stop codons were observed, and canonical donor–acceptor splice sites (GT/AG) and polyadenylation recognition sequences (AATAAA or ATTAAA) were present in all sequences. Exon sizes are similar to those found in other birds, but longer introns (in particular, introns 1 and 2; supplemental Fig. S2) resulted in total gene lengths from start to stop codons of 4,370 bp for Lasc-UAA, 3,626 bp for Lasc-UBA, and 3,513 bp for Lasc-UCA as compared to a range of 1,957–2,555 bp in galloanseriform species (Japanese quail Coja-B1 [AB078884] and goose *Anser* sp. isolate G5 [AY387655], respectively).

The region spanning intron 3 to the beginning of the 3′UTR is highly similar among all loci, with a minimum pairwise nucleotide identity of 99% (between Lasc-UBA

Fig. 1 Amino acid alignment of avian MHCI sequences and reptilian outgroup. The first residue of the α1 domain is considered as alignment position 1. *Dots* indicate identity with Lasc-UAA, while *dashes* indicate gaps. Conserved sequence features with predicted functional or structural roles are shown as: peptide mainchain-binding residues (*dark grey shading*), intra- and interdomain contacts (*light grey shading*), intradomain disulfide bridges (*horizontal lines*), N-glycosylation site (*bracket*), CD8 binding sites ("*8*"s), critical CD8 co-receptor sites (*), phosphorylated residues in cytoplasmic tail (*black diamond*) (Bjorkman et al. 1987; Saper et al. 1991; Grossberger and Parham 1992; Kaufman et al. 1994). Sequence sources are: Grca (Florida sandhill crane, *Grus canadensis pratensis*, AF033106); Coja (Japanese quail, *C. japonica* Coja-B1, AB078884); Gaga (chicken, *G. gallus* BF2, AL023516); Mega (turkey, *M. gallopavo* alpha chain 2, DQ993255); Anpl (mallard duck, *A. platyrhynchos* UAA, AY885227); Acar (great reed warbler, *Acrocephalus arundinaceus*, AJ005503); Sppu (tuatara, *Sphenodon punctatus*, DQ145788)

and -UDA, in 1,382 bp of sequence). There is a single variable site within the first 173 bp of 3′UTR sequence (of 807–857 bp in total) among all four loci. After this point, Lasc-UBA sequence cannot be aligned to those of the other three loci. In contrast, the 3′UTR sequences of Lasc-UCA and -UDA are identical over their entire lengths, and highly similar to Lasc-UAA (98.6% nucleotide identity), excepting a 47-bp deletion occurring 72 bp before the end of the 3′UTR. The first and second introns have much lower levels of nucleotide identity among loci and regions with long insertions/deletions, with the exception of Lasc-UAA and -UBA that share 97.9% sequence identity across the entire second intron (with a total size of 752–754 bp).

Conservation of sequence features with predicted functional or structural roles

Eight residues in the α1 and α2 domains contact mainchain atoms of the bound ligand, and anchor the peptide termini in a sequence-independent manner (Bjorkman et al. 1987; Saper et al. 1991; indicated by dark grey shading in Fig. 1). Classical MHCI genes of human and mouse have no more than two amino acid replacements at these sites, with substitutions tending to be of a conservative nature (Shum et al. 1999). Lasc-UAA*11 shows no deviation from the consensus "YYRTKWYY" sequence (in non-mammalian vertebrates, arginine is substituted for tyrosine at alignment position 87 [human leukocyte antigen (HLA)-A2 position 84]; Kaufman et al. 1994). Lasc-UBA has a slightly disfavoured substitution of asparagine (N) for tyrosine (Y) at position 7 (Betts and Russell 2003). Lysine (K) is substituted for arginine (R) at position 148 of Lasc-UCA, in addition to the transcribed sequence from the great reed warbler shown in Fig. 1, all described alleles in the scarlet rosefinch (*Carpodacus erythrinus*; Promerová et al. 2009),

EXON 1   (LEADER PEPTIDE)

```
Lasc-UAA    MGP--GRALAR-LLLLLGVLGGAAG
Lasc-UBA    ...--....-............
Lasc-UCA    ...--....GLG.R....LF.....
Grca        .V.--.G..GLG.G..........I
Coja        ..S--WG..ELG.-..C-AVC....
Gaga        ...--CG..GLG.-..A-AVC...A
Mega        ..S--CG..GWG.-.....AVC....
Anpl        ...GRAG..GLG.G...........S
Acar        .AAAL.LGAILG..G...EP...TR
Sppu        .YL--FGP.FLLGA.CG.TR.ALQ.
```

EXON 2   (α1 DOMAIN)

```
                 10        20        30        40        50        60        70        80        90
Lasc-UAA    GPHSLRYFYVAVS--EPSPGVPQFVAVGYVDGNGIMRYDSETGRAVPRADWMAAN--LDQAYWDRETQILQSTQQVDRVNLETARSRYNQS-R
Lasc-UBA    ......N....--........V.S........EM.G......--.......GN......S...L.....-.
Lasc-UCA    ...T.M.--....YMI.....V.L.......G......--...Q..G.RNE...H....ILGV....-G
Lasc-UDA    -...H..N.Q.L-......E.EA..L...L.S.......T..G.......-...SQ...S.NGY..A.AD...V......-G
Grca        EL.....IG..--......I.V....L.S.......M.LGRLDEGT.--..Q......RG.NN..FY.CG.G.L.G....-G
Coja        EL.T...QT.MT--D.G.EL.W.FE.....EIF.H.N.TGR.Y...TE....S--M..Q...GQ.E.A..NE.K..MS.VNLAWL...T-S
Gaga        EL.T...IQT.MT--D.G..Q.W..T.....ELFVH.N.TAR.Y...TE.I..K--A..Q...GQ...G.GNE.I..E..GILQR....T-G
Mega        E......DT.MT--D.GT.L.W..S......EIFVH..TAR.Y..HTE.VK.PGAV.PE...E.C....R.E..Y..S.GILQE.F...-G
Anpl        E......T..--D.......Y....EAFTY.....R.TE..V..I....-M..Q...MV.E.F.N.E.IY.I..D.L.E....-.
Acar        VL...H..LD....--...I......M.F...IPFT.....R..ME.LTE..IKDS--A.PE....SQ...GVGS.H.YARS...L.E.....-G
Sppu        S......TG..GLD.GA.L.E.M....EQQLDWF...RRQSE...E...Q..-EGPQ..E.Q..N..GWKAWFH...P.L.Q.....QT
```

EXON 3   (α2 DOMAIN)

```
                 100       110       120       130       140       150       160       170       180
Lasc-UAA    GAHTVQTMYGCDLLEDGSTRGYYQDAYDGRDFIALDMDTMTFTAADVGAQITKRKWEEDGTVAERMKQYLQNTCIEWLGKYVSYGRAVLERK
Lasc-UBA    .V.........Q.......R.Y.........F..........AA....................R.............
Lasc-UCA    ....R.R.L.........N..............R.........KKQH.VK.......R..............
Lasc-UDA    R...R.R.L........I...W.Y........F......I...AV........K.E...R.HC.EI......R..............
Grca        ....L.L........W.Y.......A.F.T.......AA........A......A....N...V...T........
Coja        ES..W.R.A...I...TN.....YG...K....F.K......VPE.VP.......G-GD...Q.H..EEI.PQ..WR..EH.K.E.G.T
Gaga        .S....W.....I..G.PI.....M.......T.F.KG.....VPE.VP.......E-SEP..W.N..EE..V...RR..E..K.E.G.R
Mega        .S....R......I.GNST....N.Y........F.K.KK....VPE.VL.......A.P.YVDQR.H..EE..VQ..RRHMEH.K.E...R
Anpl        .S..R.W.......S.I..FD.HG...K..LTF.K.A..Y......QE..F......F..E.......R....KD....R
Acar        .L...LRV..E..S...V..SERFG.......SF.LESGR.M...SA.E..R.R..HE.I....QTN..KHE.P...Q..G..QKE....
Sppu        .S..L.V....E.RP....G.F..HG......FNY.AN.C.WV.PVPQ.E...T....NA.TADLQHYRS..EGE......R..LA..AES.RPR
```
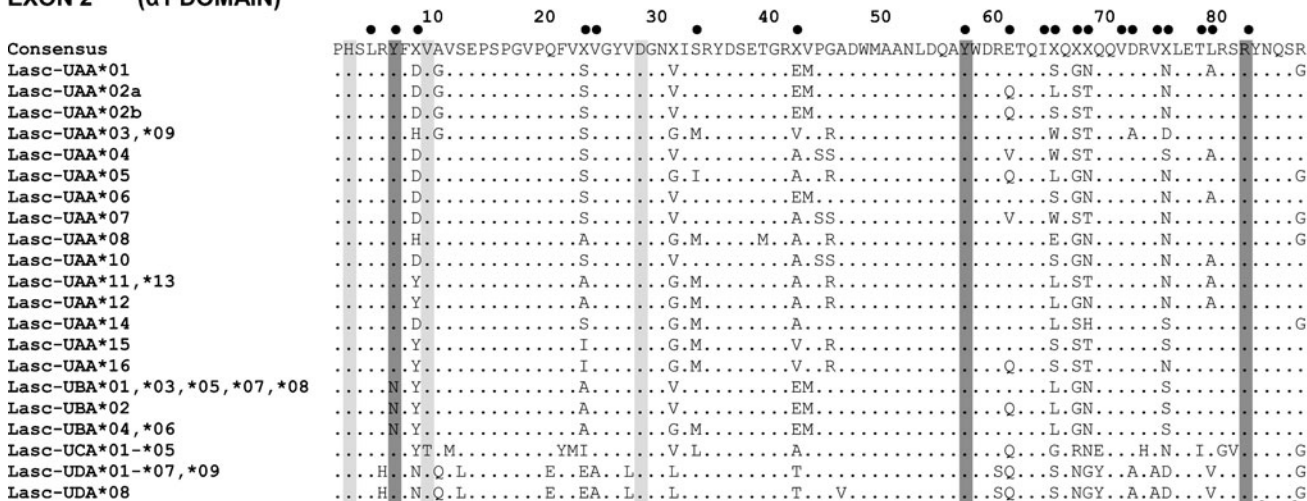
EXON 4   (α3 DOMAIN)

```
                 190       200       210       220       230       240       250       260       270
                                                     **   **  *              *
Lasc-UAA    EPPAVRVSGKEA-HGILTLRCRAHGFYPRPITVSWLKDGEVRDQETEWGSVAPNSDGTYYTWASIEARPGEQDKYRCRVEHASLPEPGLYAW
Lasc-UBA    ..........-........I.....................
Lasc-UCA    ..........-........I.....................
Lasc-UDA    ..........-........................
Grca        .R.T.....Q.T-P....H...Y...........H....R..I.......T....E.K..........T......F.
Coja        .Q.E...W....-D.....S.........A....V...A.LG.D.HS.GIV......H..VT.D.L.EDG...Q.......Q....P.
Gaga        .R.E...W....-D.....S.........V.......A.DAHS.GIV..G.....H..VT.D.Q..DG...Q..........S.
Mega        ...E...W..KT-D.....S.........D...M...VAQN.D.QS.GIT......H.LVT.D...EDG....Q........QQ...S.
Anpl        .R.E....RM..-DK....S........SI......M.QE...QR..TV.....HV..T.DVL..DW...Q........Q..FS.
Acar        ...D.H.....E-..T.I.S.H.Y...KT.A.N.M.GD.IW.......G.V......FH...R...L.E.REQ.......PGM....IF.
Sppu        .-.Q.L..DRPTPD.LTR.S.........D.A.V..RN.AAMSG..QSWGIV.SG....Q.R.T...ID.SSDAR.Q.C...E...AQDLRV..
                                    8888888888            888888888888
```

EXONS 5–8   (TRANSMEMBRANE AND CYTOPLASMIC DOMAINS)

```
                 280       290       300       310       320       330       340
                                                     ◆        ◆    ◆
Lasc-UAA    E-TESNLWAIVLGVAVAVGAV--AAICGFVVWKRKSGKKKA-GYGMASSTDSGSGSTCTGVTA---
Lasc-UBA    .-................L.--.........Q.......-...............G.....---
Lasc-UCA    .-................--...................---
Lasc-UDA    .-S..............--.....L.......T..-..................---
Grca        .-S....L......IA.IL..--I..IVG.IL.QR.......R.T..SAQVMN.--
Coja        KWPQPS.VP.IVR.VIPIVVIAIT.GI..TIYR.HA...EK-..NA.P.Q.G..S..SNQTL-
Gaga        .PPQP..VP..A....IV.IAIMVGV..IIYR.HA...GK-.NI.PDREG..S.SS..SNPAI-
Mega        .PSQP..VP..V..I..IV.IAIVGGV..IIYRSHA...EK-..NI.PAQ.G..T.SG..SDRPVI
Anpl        .-PQ...IP..V......V..-I..LA..AI..S.Q..GK-..NV.PGSEG..N.SNA.SNPSV-
Acar        .P.SGRNLTV.VA..S.ITAILILI.LT...IL..LQ..RRDRD..N..AGK.LVTNGST.ARITA--
Sppu        .-PK..VLL..GA.LG.LVVL-A..AA.IGLYI.---.RGT-..KPT.TG.RA.ENSSNENVKA-
```

EXON 2   (α1 DOMAIN)

```
                    10        20        30        40        50        60        70        80
Consensus           PHSLRYFXVAVSEPSPGVPQFVXVGYVDGNXISRYDSETGRXVPGADWMAANLDQAYWDRETQIXQXXQQVDRVXLETLRSRYNQSR
Lasc-UAA*01          ......D.G..........S......V........EM..............S.GN....N...A.....G
Lasc-UAA*02a         ......D.G..........S......V........EM..............Q...L.ST.....N........
Lasc-UAA*02b         ......D.G..........S......V........EM..............Q...S.ST.....N........
Lasc-UAA*03,*09      ......H.G..........S......G.M....V..R..............W.ST...A..D........
Lasc-UAA*04          ......D.............S......V......A.SS.............V...W.ST.....S...A.....
Lasc-UAA*05          ......D.............S......G.I....A..R.............Q...L.GN.....N........G
Lasc-UAA*06          ......D.............S......V........EM.............S.GN....N...A.....
Lasc-UAA*07          ......D.............S......V......A.SS.............V...W.ST.....N........G
Lasc-UAA*08          ......H.............A......G.M....M.A..R...........E.GN....N...A.....G
Lasc-UAA*10          ......D.............S......V......A.SS.............S.GN....N...A.....
Lasc-UAA*11,*13      ......Y.............A......G.M....A..R.............L.ST.....N..A..
Lasc-UAA*12          ......Y.............A......G.M....A..R.............L.GN....N..A..
Lasc-UAA*14          ......D.............S......G.M....A..R.............L.SH.....S........G
Lasc-UAA*15          ......Y.............I......G.M....V..R.............S.ST.....S........
Lasc-UAA*16          ......Y.............I......G.M....V..R.............Q...S.ST.....N........
Lasc-UBA*01,*03,*05,*07,*08  .....N.Y............A......V........EM.............L.GN.....S........
Lasc-UBA*02          .....N.Y............A......V........EM.............Q...L.GN.....S........
Lasc-UBA*04,*06      .....N.Y............A......G.M......EM.............L.GN.....S........
Lasc-UCA*01-*05      ......YT.M.........YMI......V.L....A...............Q...G.RNE...H.N...I.GV.....G
Lasc-UDA*01-*07,*09  ....H..N.Q.L......E..EA..L...L.......T.............SQ...S.NGY..A.AD...V.......G
Lasc-UDA*08          ....H..N.Q.L......E..EA..L...L.......T...V.........SQ...S.NGY..A.AD...V.......G
```

EXON 3   (α2 DOMAIN)

```
                    90       100       110       120       130       140       150       160       170       180
Consensus           GAHTXQRMYGCDLLEDGSTRGYXQYAYDGRDFIAFDMDTMTFTAADAXAQITKRKWEEDGTVAERXKQYLXNTCIEWLRKYVSYGRAVLERK
Lasc-UAA*01          R...R.L.................Y.S........L..........A.............QW...Q...........
Lasc-UAA*02          .V..V...................Y.N........L..........VG...........AE...W...K......G......
Lasc-UAA*03          ....W.......Q...........W...E........G........VA...........M...Q..........
Lasc-UAA*04          ....R...................Y.N........L..........VG...........W...K........
Lasc-UAA*05          ....R...................Y.N........L..........G...........QL...Q........
Lasc-UAA*06          ....W...........N.......Y........L.GM.........G...........M...Q.......G.
Lasc-UAA*07          R...W.......Q...........W...E........G........A...........M...N........
Lasc-UAA*08          R...R...L...............Y.N........L..........A...........QL...Q........
Lasc-UAA*09          ....W.......Q...........W...E........G........A...........M...Q........
Lasc-UAA*10          ....R...................Y.N........L..........VG...........AE...K......
Lasc-UAA*11          ....V.T.................Y.D........L..........VG...........M...G.......
Lasc-UAA*12          .V..V.T.....Q...........Y.N........L..........VG...........AE...W...K......G.......
Lasc-UAA*13          ....V.T.....Q...........Y.D........L..........VG...........M...Q......G.......
Lasc-UAA*14          R...W.T.....Q...........Y........L.GM.........G...........M...K........
Lasc-UAA*15          ....W...................Y.D........L..........VG...R......AE...W...K........
Lasc-UAA*16          ....R...................Y.N........L..........VG...........QL...Q.......
Lasc-UBA*01,*05,*07  .V..V.T.....Q...........R.....................A...........M...Q........
Lasc-UBA*02          .V..V.T.....Q........R..T.....................A...........M...Q........
Lasc-UBA*03          .V..V.V................S.D.......T............A....R.......R...N........
Lasc-UBA*04          .V..V.T.....Q...........R.....................A...........QL...Q........
Lasc-UBA*06          .V..V.T.....Q...........R.....................A...........W...Q........
Lasc-UBA*08          .V..V.T.....Q...........R.....................V...........M...Q........
Lasc-UCA*01,*05      ....R...L...............Y.N........L..........VG....R......KKQH.VK........
Lasc-UCA*02          ....R...L...Q...........Y.N........L..........VG....R......KKQH.VK........
Lasc-UCA*03          ....R...L...............Y.N........L..........G....R......KKQH.VK......G...
Lasc-UCA*04          ....R...L...............Y.N........L..........G....R......KKQH.VK........
Lasc-UDA*01          R...R...L...............W.............I...V...........K.E...R.HC.EI.......N........
Lasc-UDA*02          R...R...L........I......W............IT...V...S.......K.E...R.HC.EI...........
Lasc-UDA*03,*05,*07  R...R...L........I......W.............I...V...........K.E...R.HC.EI...........
Lasc-UDA*04          R...R...L........I......W.............I...V...........K.E...R.HC.EI...........
Lasc-UDA*06          R...R...L........I......W.............I...V...........K.E...R.HC.EI.......N........
Lasc-UDA*08          R...R...L........I......W............I...V...R........K.E...R.HC.EI.......N........
Lasc-UDA*09          R...W...L........I......W.............I...V...........K.E...R.HC.EI.......N........
```
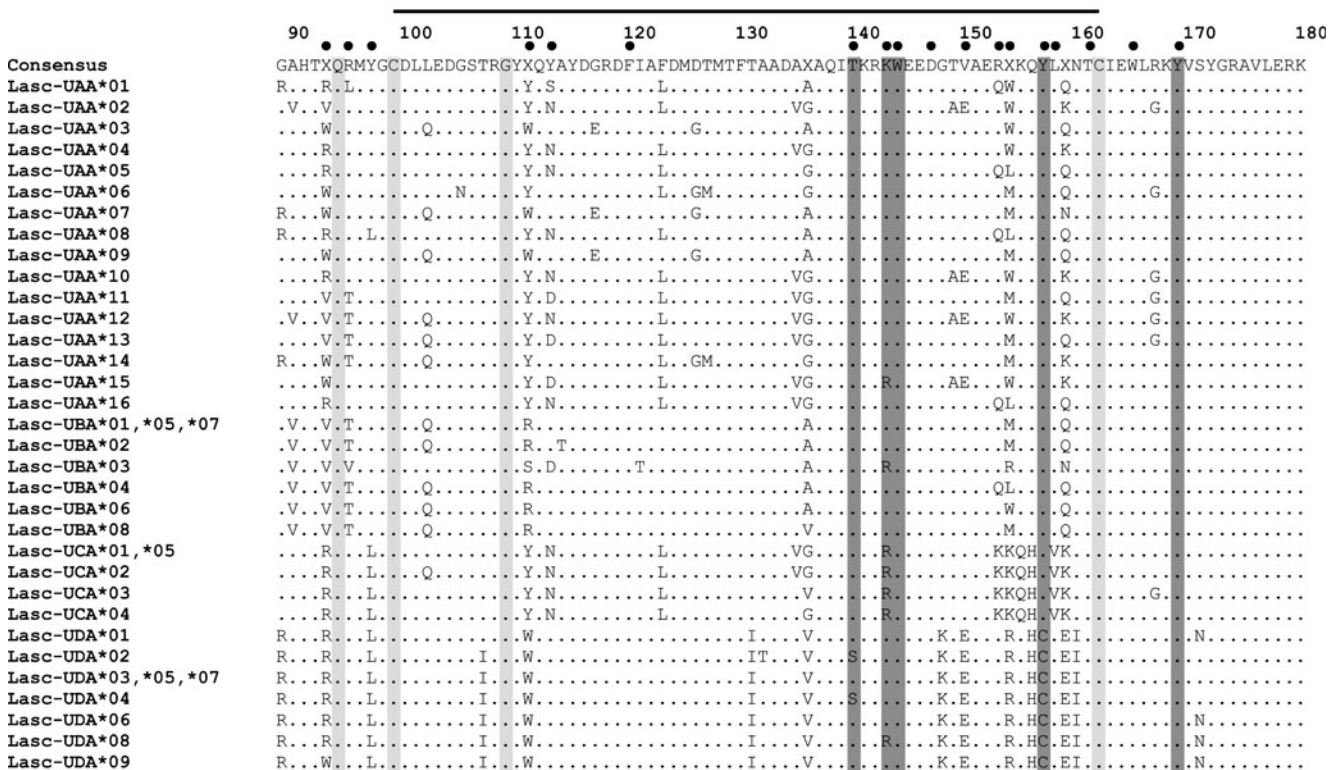
Fig. 2 Alignment of translated amino acid sequences for PBR exons of red-billed gull MHCI genes. Alleles with identical translations are shown as single sequences. *Dots* indicate identity to the majority rule consensus sequence. Annotation of conserved residues follows Fig. 1, and codons corresponding to PBR residues in human HLA-A2 are indicated by *black circle*. (Bjorkman et al. 1987; Saper et al. 1991)

both nonclassical (YF1, formerly YFV) and classical (BF2) chicken MHCI genes (Afanassieff et al. 2001), and red-billed gull alleles Lasc-UAA*15, Lasc-UBA*03, and Lasc-UDA*08 (identified during RSCA screening, and shown in Fig. 2). In Lasc-UDA, the tyrosine (Y) at position 162 is replaced by cysteine (C). Y→C is a disfavoured substitution type in extracellular proteins (Betts and Russell 2003),

and results in an unpaired cysteine within the α2 domain of all Lasc-UDA alleles (Fig. 2, position 157). A conservative substitution of serine (S) for threonine (T) also occurs in alleles Lasc-UDA*02 and Lasc-UDA*04 (Fig. 2, position 140) as well as classical MHCI genes of the chicken (Livant et al. 2004) and scarlet rosefinch allele CaerU*23 (Promerová et al. 2009). Thus, no red-billed gull allele has

more than two substitutions from the consensus sequence of mainchain-binding residues, although some nonconservative amino acid replacements are observed.

Of 18 residues involved in MHCI intra- and interdomain contacts that are highly conserved across vertebrates (Grossberger and Parham 1992), 17 are strictly conserved in all red-billed gull genes (Fig. 1). Threonine (T) at position 10 is also conserved in Lasc-UCA, but is replaced by valine (V) in Lasc-UAA, -UBA, and -UDA. The T→V

substitution is considered neutral in extracellular proteins (Betts and Russell 2003) and is also seen in the transcribed great reed warbler sequence.

Residues that form mainchain and domain contacts are also conserved in chicken allele BF2*2101, for which two conformational structures have been experimentally determined (Koch et al. 2007). Predicted PBR residues for these chicken MHCI molecules largely coincide with those of human HLA-A2 (exceptions are chicken PBR residues 61
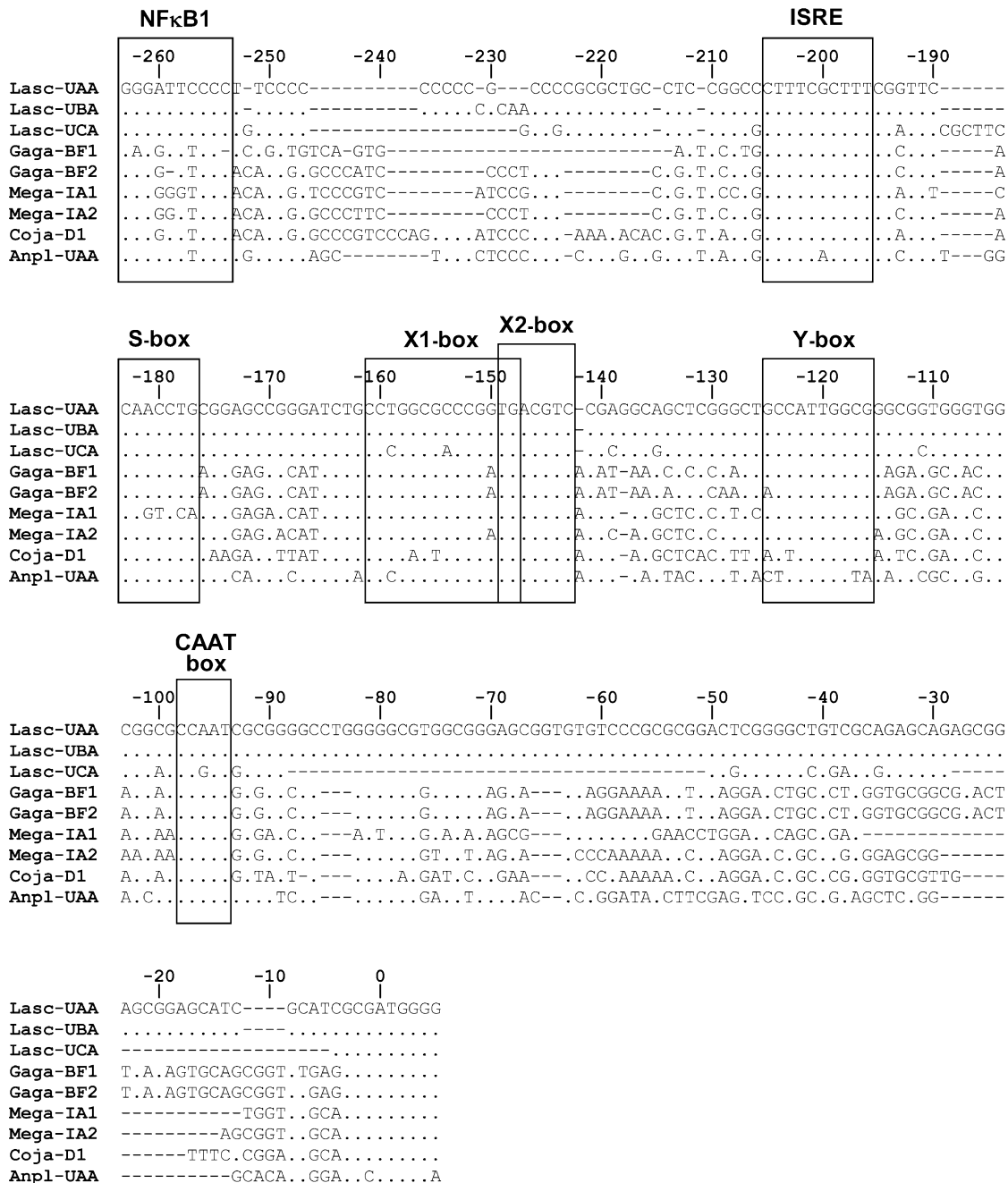
Fig. 3 Avian MHCI promoter sequences. Alignment position "0" is the first position of the start codon triplet. *Dots* indicate identity with Lasc-UAA, while *dashes* indicate gaps. Regulatory element motifs are *boxed*. Sequence sources are as in Fig. 1

and 121, which show low variability in the red-billed gull). However, no red-billed gull allele possesses the unusual combination of small residues at both positions 69 and 97, and residues of counterbalancing charge at positions 9 and 24 (basic and acidic residues, respectively; residue numbering follows Fig. 2) described for BF2*2101. It therefore seems unlikely that the enlarged central cavity of the peptide-binding groove and novel charge-transfer system that permits promiscuous binding of peptides of unrelated sequence by BF2*2101 (Koch et al. 2007), is shared by the red-billed gull.

MHCI promoter elements are conserved in the red-billed gull (Fig. 3), although this region was not sequenced for Lasc-UDA. The enhancer A NFκB1 binding site of all red-billed gull sequences is identical to those of human HLA-A and HLA-B (van den Elsen et al. 1998), whereas substitutions occur in other avian sequences. Identifiable interferon-stimulated response element and SXY module elements (van den Elsen et al. 1998) are present in all avian sequences. Deletions are seen in the proximal promoter region of Lasc-UCA, and a point mutation occurs within the CAAT-box, but it is unknown how these might affect the initiation of transcription. cDNA transcripts were isolated for each locus; thus, while these differences may modulate expression level, they do not appear to result in pseudogenization.

Sequence variation in peptide-binding region exons

Complete multilocus genotypes were obtained for 470 individuals using a combination of RSCA and sequencing of gene-specific amplification products. All patterns of allelic segregation within known family groups were consistent with Mendelian inheritance at each locus. Five reference alleles were tested for use in the RSCA procedure; of these, Lasc-UAA*03 and UAA*04 produced

the highest quality data and greatest resolution between known alleles and were used in the screening of test samples (Fig. 4, supplemental Fig. S3). Some alleles could not be resolved by RSCA typing alone (listed in caption to Fig. 4). In particular, five of the nine alleles at locus Lasc-UDA could not be differentiated, and genotype assignments were confirmed using direct sequencing of gene-specific PCR products for all samples at this locus.

Thirty-eight unique nucleotide sequences were identified from exon 3 screening and sequencing. Sequencing of exon 2 revealed one allele (Lasc-UAA*02) with two different exon 2 haplotypes (Lasc-UAA*02a and b), thus raising the total number of alleles to 39 (deposited in GenBank under accession numbers HM025950–HM025988). For both exons combined, 34 unique alleles were found at the amino acid level. When translated, no Lasc-UAA alleles shared identical amino acid sequences across both exons, whereas two to three alleles identical at the amino acid level were found in each of the other three genes (Table 2, Fig. 2). Lasc-UAA has the greatest number of alleles and greatest nucleotide and amino acid variability among alleles (Table 2). Histograms of allele frequency data (Fig. 5) show a much more even distribution for the 16 Lasc-UAA alleles than is found for the other three genes, and the observed homozygosity at this locus is significantly lower than expected under neutrality, consistent with the action of balancing selection (Table 3).

Thirty of the 38 PBR residues have predicted roles in sequence-dependent peptide binding—of these, 22 (73.3%) show non-overlapping amino acid identities between at least two genes (Fig. 2, Supplemental Fig. S4). These differences are mostly seen in comparisons involving either Lasc-UCA or -UDA, which have multiple PBR residues that are both invariant and locus-specific. Lasc-UCA is invariant at all PBR residues, while Lasc-UDA varies at a single non
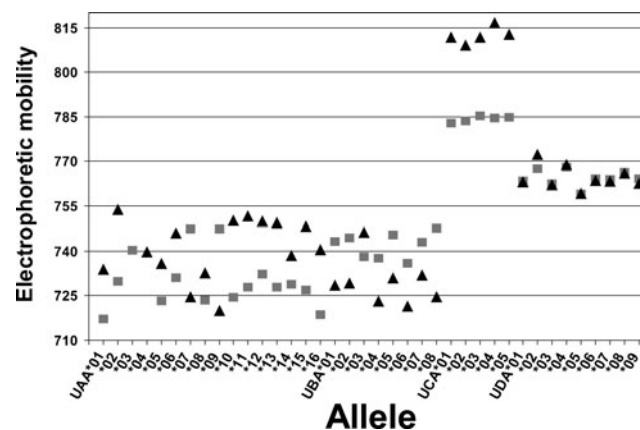

Fig. 4 Mean electrophoretic mobility of alleles hybridized with FLR-UAA*03 (*black triangle*) and FLR-UAA*04 (■). Alleles not distinguished by RSCA alone are: UAA*16 vs. a combination of UAA*04 and UAA*16; UAA*07 vs. UBA*08; UDA*01, *03, *06, *07, and *09 from each other; and alleles at the upper size range of UBA*01 from those at the lower size range of UBA*02. Mean values for these alleles were calculated following final genotype assignment by direct sequencing of PCR products. Standard deviation error bars do not extend beyond value symbols

Table 2 Summary of sequence variation and global tests of positive selection for PBR exons

	Sequence polymorphism				Positive selection				
	N_A	S_{nt}	S_{AA}	π	d_N	d_S	d_N/d_S	Z	p
All sites[a]									
Lasc-UAA	17	79	40	0.056±0.006	0.054±0.009	0.062±0.013	0.87	−0.562	1
Lasc-UBA	8 (6)	29	14	0.017±0.003	0.013±0.004	0.029±0.009	0.45	−1.749	1
Lasc-UCA	5 (4)	7	4	0.006±0.002	0.004±0.002	0.010±0.005	0.4	−1.087	1
Lasc-UDA	9 (7)	10	7	0.006±0.002	0.005±0.002	0.011±0.007	0.46	−0.826	1
PBR sites[a]									
Lasc-UAA	17	41	19	0.204±0.030	0.228±0.046	0.136±0.048	1.68	1.663	0.049
Lasc-UBA	5	13	7	0.049±0.013	0.051±0.022	0.041±0.025	1.24	0.360	0.360
Lasc-UCA	1	0	0	0	0	0	–	–	–
Lasc-UDA	3 (2)	2	1	0.005±0.003	0.003±0.003	0.010±0.011	0.30	−0.518	1

[a] Total size 540 bp (179 residues) for all sites, 90 bp (30 residues) for PBR sites excluding mainchain-binding sites

N_A number of alleles (in brackets, number of unique amino acid translations if different from N_A), S number of variable nucleotide (nt) or amino acid (AA) sites; π nucleotide diversity ± SE, d average rate of nonsynonymous (d_N) or synonymous (d_S) substitutions per site ± SE, Z p results of one-tailed Z tests of strict neutrality (H$_o$, $d_N=d_S$) versus positive selection (H$_a$, $d_N>d_S$)

mainchain-binding site. In contrast, 29 of 30 non mainchain-binding residues show a full or partial overlap in amino acid identity between Lasc-UAA and -UBA, with Lasc-UBA alleles tending to encompass a subset of the variation found in Lasc-UAA.

All measures indicate substantial intralocus recombination within Lasc-UAA (Table 4). Notably, the per-site population recombination rate (ρ) slightly exceeds the

estimate of mutation within the population (θ), indicating that both mutation and recombination contribute substantially to total sequence variation within this locus. In contrast, the other three genes show little or no evidence of intralocus recombination as a major force acting to generate allelic diversity.

Lasc-UAA and -UBA show evidence of four significant interlocus gene conversion events and four additional,
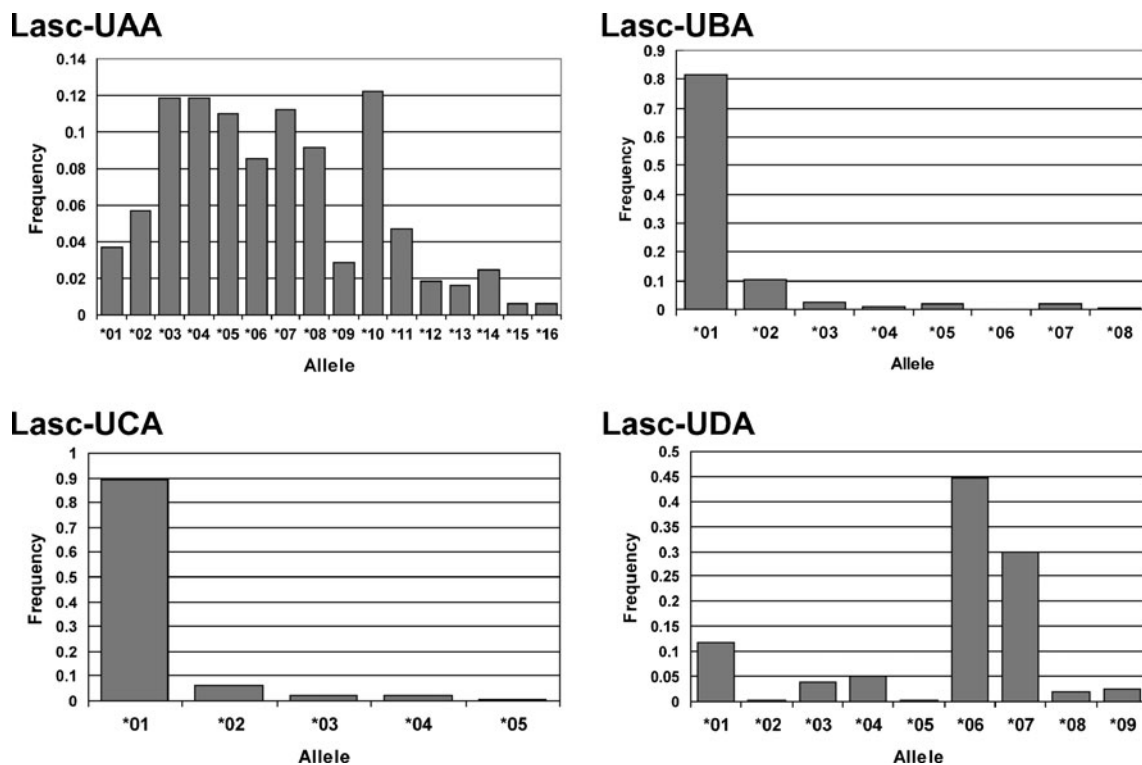
Fig. 5 Allele frequencies of red-billed gull MHCI genes

Table 3 Ewens–Watterson neutrality tests of allele frequencies

	$F_{(obs)}$	$F_{(exp)}$	p
Lasc-UAA	0.093	0.237	<0.001
Lasc-UBA	0.683	0.424	0.917
Lasc-UCA	0.799	0.575	0.836
Lasc-UDA	0.309	0.424	0.265

$F_{(obs)}$ observed homozygosity, $F_{(exp)}$ expected homozygosity under selective neutrality, calculated from 20,000 simulated samples

although statistically unsupported, regions of sequence identity within PBR exons 2 and 3 (supplemental Fig. S5). Conversely, Lasc-UCA and -UDA have numerous locus-specific substitutions throughout both PBR exons, although regions of nucleotide sequence identity that are possibly indicative of gene conversion cover the 5′ end of exon 3.

These differences in recombination and/or gene conversion are evident in the network reconstruction of PBR exon sequences, where Lasc-UCA and -UDA form well-differentiated locus-specific clades (Fig. 6). Alleles of Lasc-UBA also cluster in a gene-specific manner, although shorter branch lengths and multiple parallel splits indicate greater conflict in the phylogenetic signal differentiating this clade from alleles of Lasc-UAA. Lasc-UAA does not form a single gene cluster, and numerous alternative splits indicate the prevalence of conflicting evolutionary signals among alleles of this locus, consistent with the much higher estimate of intralocus recombination.

Inference of positive selection

The ratio of nonsynonymous to synonymous substitution ($\omega=d_N/d_S$) is a widely used measure of the action of selection on coding sequences, with $\omega<1$ indicative of negative (purifying) selection and $\omega>1$ serving as a signal of positive (diversifying) selection. Global estimates of selection across predicted PBR sites (excluding those

Table 4 Summary of intralocus recombination analyses for concatenated exon 2 and 3 sequences

	GARD	Rm	$4N_e r$	ρ	θ
Lasc-UAA	3	19	17	0.032	0.028
Lasc-UBA	0	1	0	0	0.010
Lasc-UCA	0	0	0	0	0.003
Lasc-UDA	0	0	1	0.002	0.004

GARD (Kosakovski Pond et al. 2006) breakpoints with $p<0.05$

Rm minimum number of recombination events, $4N_e r$ recombination parameter, ρ population-scaled recombination rate, θ Watterson estimate of the mean population mutation rate

involved in binding peptide mainchain atoms) suggest that only Lasc-UAA is subject to diversification resulting from positive selection (Table 2).

Estimation of the spatial variation in selection can be more biologically meaningful than global tests when positively selected sites are interspersed with residues under strong functional constraint or when a priori restriction of analyses to sites "known" to be likely targets of positive selection (e.g., PBR residues) is based upon multiple sequence alignment rather than experimental determination of protein structure in the focal species (Yang et al. 2000; Hughes and Friedman 2008). OmegaMap analyses indicate that Lasc-UAA has the greatest number of residues with posterior probabilities of positive selection ≥0.95 (Fig. 7), and that a large proportion of these residues have predicted roles in sequence-dependent peptide binding (17 of 24 sites, or 70.8%).

Unlike Lasc-UAA, positive selection does not appear to greatly promote allelic diversification at the other MHCI loci. Six predicted PBR residues of Lasc-UBA showed evidence of positive selection at the 0.95 posterior probability cutoff. However, variation at three of these sites largely coincides with inferred gene conversion tracts between Lasc-UAA and -UBA. It is possible that the observed pattern at these sites reflects an absence of negative selection acting upon these shared PBR residues rather than positive selection, although this interpretation assumes a directionality to gene conversion events that requires further investigation. Lasc-UCA has no PBR sites showing evidence of positive selection, while only a single residue identified for Lasc-UDA has predicted involvement in non-mainchain peptide binding. However, caution should be used in interpreting an excess of nonsynonymous substitution as evidence of positive selection for these loci as, given the low number of observed substitutions, this pattern could equally reflect stochastic variation in d_N and d_S across codons (Hughes and Friedman 2008).

Discussion

Characterization of red-billed gull MHCI genes

The four MHCI genes presented here are the first complete avian MHCI sequences described outside galloanseriforms (gamebirds + waterfowl), and the first MHC class I genes isolated from Charadriiformes (shorebirds, gulls, and allies). It is likely that these genes represent all intact MHCI gene family members of the red-billed gull, as isolation and screening procedures used three different primer pairs placed in highly conserved regions and targeting different sequence segments. However, additional pseudogenes could exist as
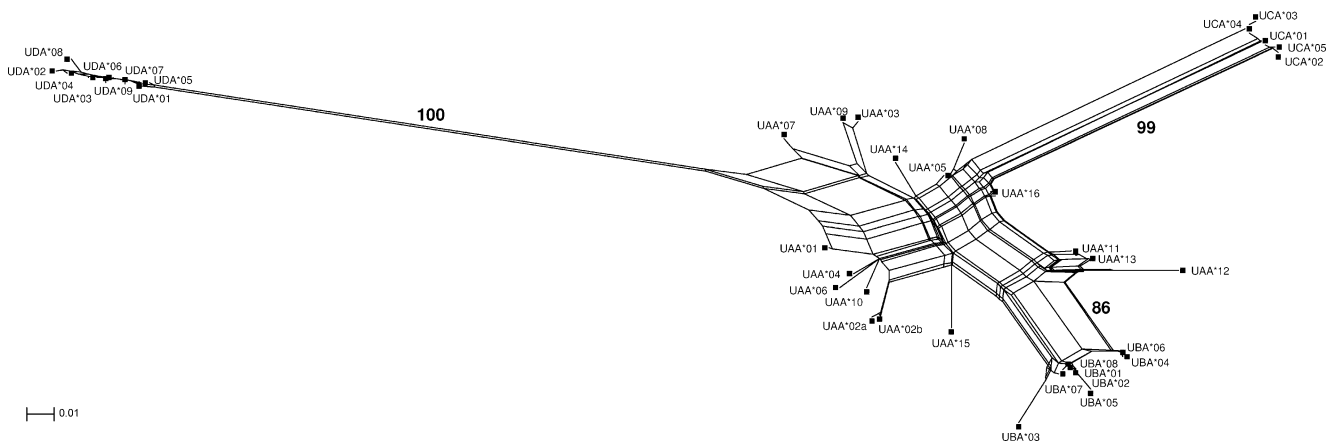
Fig. 6 Phylogenetic network of relationships among red-billed gull MHCI alleles from concatenated exon 2 and 3 nucleotide sequences. Percentage bootstrap supports are shown for Lasc-UBA, -UCA, and -UDA locus-specific clades

gene fragments or orphan exons, as are seen within the Japanese quail B locus (Coja-F, -G, and -H; Shiina et al. 2004). Sequence features typical of antigen-presenting MHCI genes are largely conserved in each red-billed gull gene, and no frameshift mutations, premature stop codons, or inappropriate splice signals were observed. Additionally, cDNA sequences were isolated for each locus, indicating that all are transcriptionally active in spleen tissue.

Only Lasc-UAA transcripts were isolated during hybridization-based cDNA library screening, whereas transcripts for the other three loci were instead obtained during PCR-based screening using gene-specific primers. The hybridization probe was amplified from a conserved region of exon 4 with a single variable nucleotide site among all loci, and was therefore unlikely to result in the biased detection of Lasc-UAA transcripts. Instead, the results suggest that only Lasc-UAA was expressed at a high enough level to produce strong positive signals from cDNA pools (i.e., positive pools likely contained multiple transcripts). In support of this reasoning, 12 of the 77 positive pools contained mixed sequences entirely attributable to transcripts of both Lasc-UAA*08 and Lasc-UAA*11. Major class I genes with high sequence variation and high expression levels across tissue types are described for the chicken and duck, and in each instance are located adjacent to the antigen transporter gene TAP2 (Kaufman 1999; Kaufman et al. 1999a; Mesa et al. 2004). Tight MHCI–TAP2 linkage may result in the coevolution of coordinated functional specificities of haplotypic alleles for these genes, and could explain the occurrence of a single, predominantly expressed MHCI locus (Kaufman 1999; Kaufman et al. 1999a). A similar position and orientation relative to TAP2 was found for Lasc-UAA from cosmid clone sequencing results (A. Cloutier, unpublished data), and Lasc-UAA might therefore represent a major class I gene in the red-billed gull.

In addition to its location relative to TAP2 and inferred higher expression level, the polymorphism and pattern of positive selection of Lasc-UAA are also consistent with its designation as a major classical MHCI gene. Among the four genes, Lasc-UAA has not only the greatest number of alleles and highest amino acid divergence among alleles, but also a more even allele frequency distribution that is suggestive of balancing selection acting to maintain allelic polymorphism, and an excess of nonsynonymous substitution concentrated in residues with predicted involvement in sequence-dependent peptide binding, which is also consistent with balancing selection for PBR diversity acting over longer evolutionary timescales (and resulting in a pattern of positive selection for individual codons, Garrigan and Hedrick 2003; Piertney and Oliver 2006). Both point mutation and intralocus recombination contribute to the observed variation of Lasc-UAA, and the effect of intralocus recombination is evident as multiple parallel splits among alleles of this gene in the network reconstruction of MHCI relationships. High levels of recombination are associated with high exon 3 sequence diversity in the scarlet rosefinch (*C. erythrinus*), although locus affiliations were unknown in this study (Promerová et al. 2009). Intralocus recombination contributes substantially to the generation of allelic diversity in the Eurasian kestrel (*Falco tinnunculus*; Alcaide et al. 2009), and to a lesser extent, is also reported for alleles of the major MHCI gene in the chicken (BF2; Hunt et al. 1994; Hunt and Fulton 1998). The relatively large size of Lasc-UAA intron 2 (754 bp) suggests that additional allelic variation could result from recombination within this region. While variation in exon 3 was screened for all individuals, exon 2 sequences were instead obtained from direct sequencing of PCR fragments in a subset of individuals, and the importance of recombination within intron 2 to total sequence variation at this locus remains to be determined.

In the chicken and duck, the occurrence of a major MHCI gene is accompanied by mutational events acting to reduce or eliminate the expression of additional loci, including
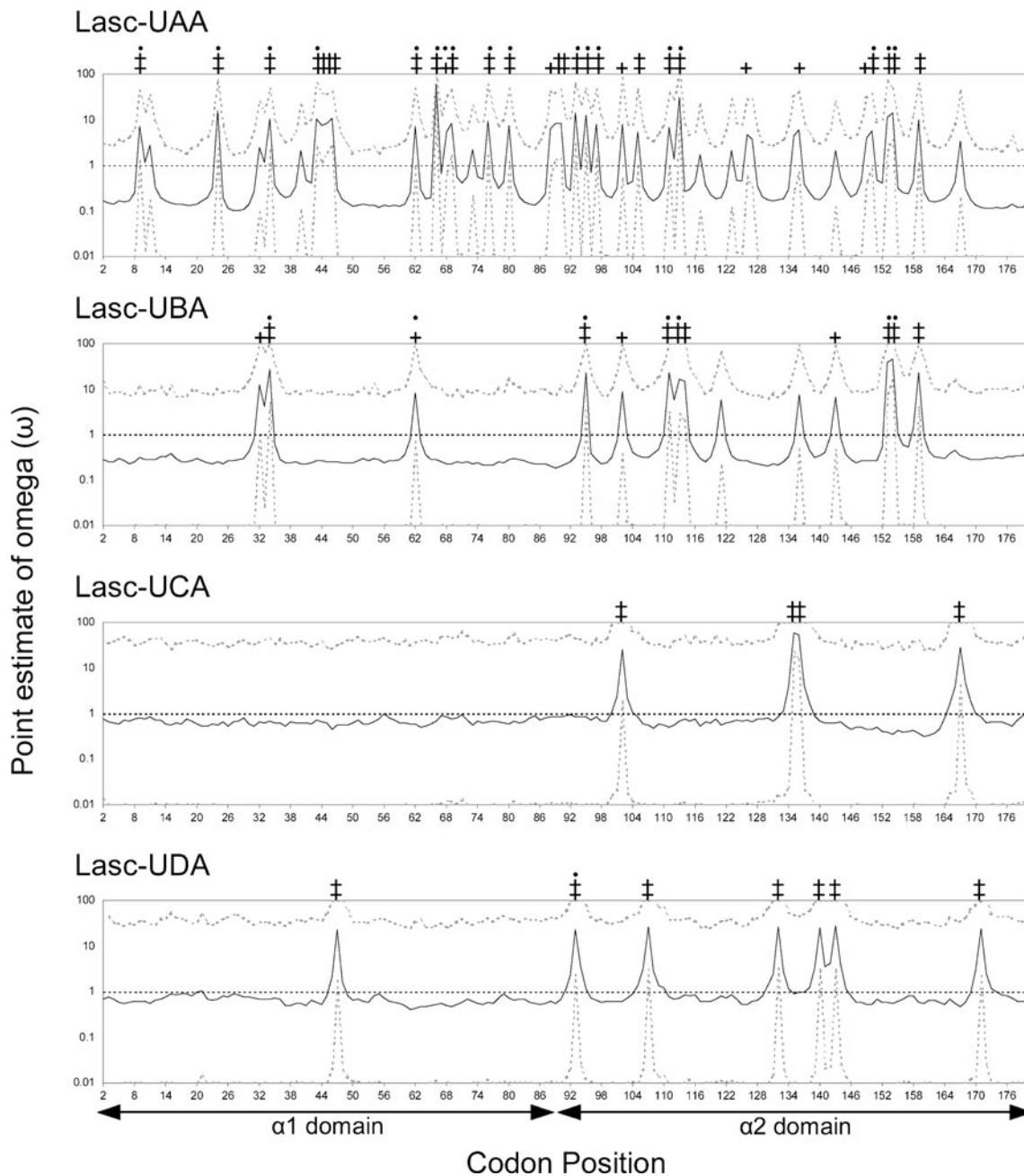
Fig. 7 Spatial variation in selection across PBR exons. Point estimates of omega (ω) are plotted as *solid black lines* on a logarithmic scale, and higher and lower 95% highest posterior probability densities as *dotted grey lines*. The selection cutoff $d_N/d_S=1$ is shown as a *dotted black line*. Sites with a posterior probability of positive selection ≥0.9 are indicated by *plus sign*, those ≥0.95 by *double dagger sign*, and predicted sequence-dependent PBR residues by *black circle*. Codon numbering corresponds to Fig. 2

deletions or nucleotide substitutions in upstream regulatory elements, missing polyadenylation signals, or premature stop codons (Moon et al. 2005; Shaw et al. 2007). In contrast, Lasc-UBA, -UCA, and -UDA each show conservation of sequence features that are typical of functional MHCI genes, and all are transcribed in the spleen. However, differences in polymorphism and selection from those seen for Lasc-UAA suggest that they function as either minor classical class Ia or nonclassical class Ib genes.

Although alleles of Lasc-UBA form a locus-specific clade in network reconstructions of PBR exon nucleotide sequences, substantial overlap occurs between Lasc-UAA and -UBA in the identities of residues with predicted involvement in sequence-dependent peptide binding, with alleles of Lasc-UBA tending to encompass a subset of the variation observed for Lasc-UAA. While this similarity suggests both loci could bind and present foreign peptides in a similar manner, the presumably lower expression level

of Lasc-UBA in spleen tissue, in combination with its more limited variation, suggests that this gene is a minor classical class I locus. Both Lasc-UBA and the minor MHCI locus of the chicken (BF1) have fewer alleles, lower sequence variation, and less evidence of positive selection than their "major" classical counterparts, Lasc-UAA and BF2, respectively (Livant et al. 2004; Shaw et al. 2007). However, the eight alleles of Lasc-UBA are less variable than BF1, with only 14 polymorphic amino acid residues observed within the α1 and α2 domains, as compared to 32 variable sites in nine alleles for BF1 (Livant et al. 2004). Furthermore, 16 of these variable BF1 sites have amino acid identities that are not entirely shared with alleles of the major BF2 locus, including a cluster of conserved locus-specific residues within the α1 domain (Livant et al. 2004). In contrast, only three positions in Lasc-UBA encode residues that are not found in any Lasc-UAA allele. Further differences are seen in the upstream regulatory regions, where BF1 alleles have deletions within the enhancer A element or transcriptional start sites of the proximal promoter (Shaw et al. 2007), whereas Lasc-UAA and -UBA share high sequence similarity across the entire promoter region and are identical at all predicted transcription factor binding sites.

Linkage relationships of putative major and minor genes also appear to differ between the red-billed gull and chicken. Serotyping and restriction fragment length polymorphism analyses of offspring from experimental matings suggest that recombination is rare within the BF (class I)–BL (class II) region of the chicken B locus (Skjødt et al. 1985; Hála et al. 1988; Kaufman et al. 1999b), although the identification of a recombination breakpoint within the BF region from DNA sequencing indicates that reshuffling of major and minor MHCI alleles among haplotypes does occasionally occur (Hosomichi et al. 2008). In contrast, although low offspring numbers per family and low polymorphism of Lasc-UBA limited analyses, we detected recombination of Lasc-UAA/UBA parental haplotypes in 7 of 12 red-billed gull family groups, suggesting that haplotypes undergo frequent recombination. An intriguing possibility is that limited Lasc-UBA variation could signal a functional constraint for alleles to retain PBR motifs that are at least somewhat compatible with all TAP translocation specificities, as recombination would frequently reassort alleles of these genes among haplotypes. Alternatively, greater sequence similarity of these red-billed gull loci could reflect a more recent gene duplication event, heightened by sequence homogenization arising from multiple gene conversion events. Fuller knowledge of the function and evolutionary relationships of these loci requires a survey of TAP polymorphism and investigation of major MHCI–TAP coevolution in the red-billed gull, quantitative measure-ment of gene expression across tissues, identification of epitopes bound by alleles of each gene, and investigation of MHCI genes in other shorebirds to infer the timing of duplication events.

In contrast to the greater sequence variation of the putative classical loci, Lasc-UCA and -UDA showed extremely limited variation. Of the few variable sites, none (Lasc-UCA) or one (Lasc-UDA) corresponded to non-synonymous substitution of residues with predicted involvement in sequence-dependent binding of ligands, and both genes have a number of locus-specific residues at these sites. Thus, both Lasc-UCA and -UDA possess divergent and mostly invariant binding-groove motifs, while largely maintaining other conserved sequence features of functional MHCI molecules, which suggest that they could be nonclassical MHC Ib genes with restricted peptide-binding specificities, or could present an alternative repertoire of non-peptidic ligands.

Nonclassical MHCI genes with limited ranges of peptide-binding capabilities or differences in the manner of contact made with the bound peptide have been identified in both humans and mice (Shawar et al. 1994; Rodgers and Cook 2005). Afanassieff et al. (2001) suggested that YF1 (formerly YFV), a nonclassical MHCI gene of the chicken Y locus, could also bind peptides in a novel manner due to substitutions at four mainchain-binding positions. More recently, Hee et al. (2010) determined the crystal structure of allele YF1*7.1 complexed with β2-microglobulin. They suggested that the "hybrid" structure of YF1*7.1, consisting of a hydrophobic binding groove situated within a protein architecture highly similar to that of classical MHCI molecules, could allow the binding of non-peptidic ligands, and demonstrated that YF1*7.1 can indeed bind nonself lipids in vitro (Hee et al. 2010). Whether nonclassical genes of the red-billed gull serve a similar function is currently unclear. Although substitutions occur at mainchain-binding sites of both Lasc-UCA and -UDA, no more than two substitutions at these sites occur in any allele and only one change involves a radical amino acid replacement. The structural importance of this radical Y→C substitution that results in an unpaired cysteine within the α2 domain of all Lasc-UDA alleles remains to be determined, although unpaired cysteines within the α1 domain are involved in MHCI homodimer formation in human HLA-B27 (Allen et al. 1999). Greater amino acid identity is seen between chicken YF1*7.1 and classical BF2*2101 than to either red-billed gull locus, whether considering the entirety of exons 2 and 3 (0.581, 0.519–0.525, and 0.491–0.513 for YF1 compared to BF2, Lasc-UCA, and -UDA, respectively), or the 27 residues identified as lining the binding groove of YF1*7.1 (Hee et al. 2010; 0.481 for YF1–BF2, versus 0.444 for YF1–Lasc-UCA, and 0.370–0.407 for alleles of

Lasc-UDA). The number of hydrophobic residues occupying these 27 sites in putative nonclassical loci of the red-billed gull (12–13 residues) is also lower not only than YF1*7.1 (16 residues) but also BF2*2101 (14 residues). Thus, while nonclassical loci of both the chicken and red-billed gull may serve to increase an otherwise limited repertoire of bound ligands presented by a major classical MHCI gene, further work is needed to determine the structure, function, and tissue expression patterns of Lasc-UCA and -UDA.

Screening of MHC variation in nonmodel organisms

Investigation of MHCI sequence evolution, and the use of these genes as molecular markers in studies of natural populations, requires an efficient and reliable methodology to screen MHCI allelic variation. RSCA allows the rapid screening of a large sample size with high reproducibility and minimal labour, and has proved effective in studies of MHC genes across a range of vertebrates (e.g., Noakes et al. 2003; Pratt et al. 2006; Worley et al. 2008; Lenz et al. 2009). RSCA is particularly suited to the initial characterization of MHCI variation in nonmodel organisms by targeting a subset of individuals carrying each allele for subsequent sequence determination. Despite the advantages of the RSCA screening approach, and the ability of our protocol to discriminate between some alleles that differed by a single nucleotide, not all red-billed gull MHCI alleles could be distinguished by RSCA alone. Further optimization will require the testing of additional reference alleles, including cloned alleles that incorporate sequence artefacts, to develop a reference panel that can unambiguously identify all alleles.

Consistent with other studies of MHCI sequence variation in birds and reptiles (e.g., Miller et al. 2007; Shaw et al. 2007; Alcaide et al. 2009; Promerová et al. 2009), sequences immediately flanking the PBR exons were highly conserved, and the placement of PCR primers within these regions resulted in simultaneous amplification of all red-billed gull MHCI genes. However, differences in intron 2 and/or the 3′UTR permitted the design of gene-specific primer pairs spanning exon 3 for each locus. By employing direct sequencing of gene-specific amplification products, it was possible to fully resolve all genotypes, to easily obtain sequences of novel alleles detected during the screening procedure, and to confirm the accuracy of allele sequences obtained from cloning. The usefulness of noncoding regions for developing locus-specific protocols necessarily relies on the extent and nature of sequence divergence between duplicate genes, which is in turn dependent upon the time since duplication, mutation rate, extent of recombination and gene conversion, and population-level processes such as genetic drift (Innan

2009). Gene-specific amplification may be impossible for some species (e.g., Westerdahl et al. 2004; Miller et al. 2007) without the use of long-range PCR using primers in adjacent genomic regions, and larger MHCI gene numbers in passerines (Westerdahl 2007) may further complicate research endeavours within this group. Nevertheless, our results indicate that efforts to sequence noncoding regions of MHCI loci could be amply rewarded by the ability to characterize individual loci, and could not only enhance the utility of these genes as molecular markers in evolutionary ecology research, but could also greatly contribute to the study of MHC gene family evolution across the breadth of avian diversity.

References

Afanassieff M, Goto RM, Ha J, Sherman MA, Zhong LW, Auffray C, Coudert F, Zoorob R, Miller MM (2001) At least one class I gene in restriction fragment pattern-Y (*Rfp-Y*), the second *MHC* gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. J Immunol 166:3324–3333

Alcaide M, Edwards SV, Cadahía L, Negro JJ (2009) MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. Conserv Genet 10:1349–1355

Allen RL, O'Callaghan CA, McMichael AJ, Bowness P (1999) HLA-B27 can form a novel β_2-microglobulin-free heavy chain homodimer structure. J Immunol 162:5045–5048

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour 10:237–251

Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. In: Barnes MR, Gray IC (eds) Bioinformatics for geneticists. Wiley, Chichester, pp 289–316

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512–518

Dean AD, Greenwald JE (1995) Use of filtered pipet tips to elute DNA from agarose gels. Biotechniques 18:980

Edwards SV, Nusser J, Gasper J (2000) Characterization and evolution of major histocompatibility complex (MHC) genes in non-model organisms, with examples from birds. In: Baker AJ (ed) Molecular methods in ecology. Blackwell Scientific, Cambridge, pp 168–207

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Garrigan D, Hedrick PW (2003) Detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57:1707–1722

Gillingham MAF, Richardson DS, Løvlie H, Moynihan A, Worley K, Pizzari T (2009) Cryptic preference for MHC-dissimilar females in male red junglefowl, *Gallus gallus*. Proc R Soc Lond B 276:1083–1092

Grossberger D, Parham P (1992) Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. Immunogenetics 36:166–174

Hála K, Chaussé AM, Bourlet Y, Lassila O, Hasler V, Auffray C (1988) Attempt to detect recombination between *B-F* and *B-L* genes within the chicken *B* complex by serological typing, in vitro MLR, and RFLP analyses. Immunogenetics 28:433–438

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp 41:95–98

Hee CS, Gao S, Loll B, Miller MM, Uchanska-Ziegler B, Daumke O, Ziegler A (2010) Structure of a classical MHC class I molecule that binds "non-classical" ligands. PLoS Biol 8:e1000557

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. BioScience 52:423–431

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M, Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken *Mhc-B* haplotype diversity. J Immunol 181:3393–3399

Hughes AL, Friedman R (2008) Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes. Immunogenetics 60:495–506

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol 6:559–579

Hunt HD, Fulton JE (1998) Analysis of polymorphisms in the major expressed class I locus (*B-F*IV) of the chicken. Immunogenetics 47:456–467

Hunt HD, Pharr GT, Bacon LD (1994) Molecular analysis reveals MHC class I intra-locus recombination in the chicken. Immunogenetics 40:370–375

Hunt HD, Goto RM, Foster DN, Bacon LD, Miller MM (2006) At least one YMHCI molecule in the chicken is alloimmunogenic and dynamically expressed on spleen cells during development. Immunogenetics 58:297–307

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Innan H (2009) Population genetic models of duplicated genes. Genetica 137:19–37

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

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

Kaufman J, Salomonsen J, Flajnik MF (1994) Evolutionary conservation of MHC class I and class II molecules—different yet the same. Semin Immunol 6:411–424

Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, Salomonsen J (1999a) Gene organization determines evolution of function in the chicken MHC. Immunol Rev 167:101–117

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

Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny H-J, 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 27:885–899

Kosakovski Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—implications for evolutionary analysis. Gene 427:117–123

Lenz TL, Eizaguirre C, Becker S, Reusch TBH (2009) RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. BMC Evol Biol 9:57

Livant EJ, Brigati JR, Ewald SJ (2004) Diversity and locus specificity of chicken MHC B class I sequences. Anim Genet 35:18–27

Longeri M, Zanotti M, Damiani G (2002) Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. Eur J Immunogenet 29:517–523

McVean G, Auton A (2007) LDhat 2.1: a package for the population genetic analysis of recombination. Department of Statistics, Oxford University. http://www.stats.ox.ac.uk/~mcvean/Ldhat/manual.pdf. Accessed 24 October 2010

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 56:192–203

Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N, Briles WE (1994) Two *Mhc* class I and two *Mhc* class II genes map to the chicken *Rfp-Y* system outside the *B* complex. Proc Natl Acad Sci USA 91:4397–4401

Miller HC, Andrews-Cookson M, Daugherty CH (2007) Two patterns of variation among MHC class I loci in tuatara (*Sphenodon punctatus*). J Heredity 98:666–677

Mills JA (1989) Red-billed gull. In: Newton I (ed) Lifetime reproduction in birds. Academic Press Ltd., London, pp 387–404

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

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Noakes MA, Reimer T, Phillips RB (2003) Genotypic characterization of an MHC class II locus in lake trout (*Salvelinus namaycush*) from Lake Superior by single-stranded conformational polymorphism analysis and reference strand-mediated conformational analysis. Mar Biotechnol 5:270–278

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity 96:7–21

Pratt BF, O'Connor DH, Lafont BAP, Mankowski JL, Fernandez CS, Triastuti R, Brooks AG, Kent SJ, Smith MZ (2006) MHC class I allele frequencies in pigtail macaques of diverse origin. Immunogenetics 58:995–1001

Promerová M, Albrecht T, Bryja J (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the scarlet rosefinch (*Carpodacus erythrinus*). Immunogenetics 61:451–461

Rodgers JR, Cook RG (2005) MHC class Ib molecules bridge innate and acquired immunity. Nat Rev Immunol 5:459–471

Saper MA, Bjorkman PJ, Wiley DC (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6Å resolution. J Mol Biol 219:277–319

Sawyer SA (1999) GENECONV: a computer package for the statistical detection of gene conversion. Department of Mathematics, Washington University in St. Louis. http://www.math.wustl.edu/~sawyer/geneconv/gconvdoc.html. Accessed 24 October 2010

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 178:5744–5752

Shawar SM, Vyas JM, Rodgers JR, Rich RR (1994) Antigen presentation by major histocompatibility complex class I-b molecules. Annu Rev Immunol 12:839–880

Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, Hanzawa K, Beck S, Kulski JK, Inoko H (2004) Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol 172:6751–6763

Shiina T, Hosomichi K, Hanzawa K (2006) Comparative genomics of the poultry major histocompatibility complex. Anim Sci J 77:151–162

Shum BP, Rajalingam R, Magor KE, Azumi K, Carr WH, Dixon B, Stet RJM, Adkison MA, Hedrick RP, Parham P (1999) A divergent non-classical class I gene conserved in salmonids. Immunogenetics 49:479–490

Skjødt K, Koch C, Crone M, Simonsen M (1985) Analysis of chickens for recombination within the MHC (B-complex). Tissue Antigens 25:278–282

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tao N, Richardson R, Bruno W, Kuiken C (2009) FindModel documentation. http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html. Accessed 24 October 2010

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

van den Elsen PJ, Gobin SJP, van Eggermond MCAJ, Peijnenburg A (1998) Regulation of MHC class I and II gene transcription: differences and similarities. Immunogenetics 48:208–221

Watterson GA (1978) The homozygosity test of neutrality. Genetics 88:405–417

Westerdahl H (2007) Passerine MHC: genetic variation and disease resistance in the wild. J Ornithol 148:S469–S477

Westerdahl H, Wittzell H, von Schantz T, Bensch S (2004) MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. Heredity 92:534–542

Wilson DJ, McVean G (2006) Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172:1411–1425

Worley K, Gillingham M, Jensen P, Kennedy LJ, Pizzari T, Kaufman J, Richardson DS (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenetics 60:233–247

Yang Z, Nielsen R, Goldman N, Pederson AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449