Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure

Li-Cheng Chen[†], Hong Lan[†], Li Sun[†], Yan-Li Deng, Ke-Yi Tang, and Qiu-Hong Wan*

Table S1. Primers used for screening BACs

Region	Primer name	Location	Primer sequence (5′→3′)	Product length (bp)	Positive BAC clones [a]
BG	N-BG	*BG*	F: AGCGTCTTGCACAGTGCAGACGT R: CGTGGACGCTCGGAGCTTCGAGA	223	**110G1**, **744D1, 285F5, 515F3, 1007B2, 1228H11,** H1-8B2F1, etc.
	110G1-F	end-primer	F: CTGGGCAGGAGGTGGTGA R: TTCTCGGAGGGCGTGTAG	404	**110G1**, **1279H12, 302C12, 844G6, 980G4, 1314A1, 922G5**, M64-6D1G2, M109-4G12A7, etc.
Core	N-IIα	MHC-IIα	F: TCCCAGAGCCCTTTCGGTAACAAC R: GCGCCGTCCGTCACCTCCTG	442	M23-1B5D8, H28-4B11E7, etc.
	N-IIβ	MHC-IIβ	F: GCATCCACAACCGGGAGCAG R: GTAGTTGTGGCGGCAGTACGTGTC	161	M23-1B5D8, H28-4B11E7, **542D2**, etc.
	N-DMA	*DMA*	F: CGACCGGTGACGTTGGGTGAAC R: CGTGGACGCTCGGAGCTTCGAGA	144	M23-1B5D8, H28-4B11E7, **514C2**, **542D2**, etc.

	N-I-1	MHC-I	F: CCCGCCTCGCCCTCACCTT R: GGGGGATTTATCAGAACGCCTACG	235	**514C2**, **542D2**, **122F9** [b], etc.
	N-Tap2	*Tap2*	F: AGTCGTGCTGCCCATCTGCT R: GCGGGGTTGTTGGAGCGATTCTAC	142	**514C2**, **1018A2**, H28-4B11E7
Class I	N-I-2	MHC I	F: GGAGGCCCAGGTGTAGTAGGTGC R: GCGGGCCGTGCTGGAGAG	288	**122F9**, **1189C6**, H24-3G7B3, H56-2A11C2, etc.
	122F9-F	end-primer	F: TGGGGTGGTGGCTCCTCCTCTT R: AGGGCCTGGTGGGTGGATGAAG	436	**122F9**, **988E8**, H24-3G7B3, H56-2A11C2

[a] Shaded BACs were used in the tiling path (Figure 1). Bold BACs are from the routine genomic library, and other BACs are from the reverse-4D library. The "etc" represents that there are more possible positive BACs in the reverse-4D library; however, we stopped screening owing to inconvenient non-gridded properties of the reverse-4D library for obtaining lots of positive BACs [S1].

[b] This BAC is a real positive clone due to universal class I primers but it does not overlap with the BACs in the Core Region.

Table S2. Comparison of the crested ibis MHC genes with those in other species

Locus [a]	No. of exons	Species [b]	Prominent features of the highest amino acid similarity	Amino acid seq. ass. no.	Positive (%)	E-value [c]	Conserved domain
Blec1	5	Rock pigeon	C-type lectin domain family 2 member B	EMC86793	87	1e-26	CLECT
Blec2	5	Rock pigeon	C-type lectin domain family 2 member B	EMC86793	87	3e-26	CLECT
zfp7	14	Chicken	Zinc finger protein 692	NP_001092826	72	0.0	COG5948 SFP1
TRIM7.2	7	Turkey	TRIM protein 7	ADU03778	97	0.0	RING BBOX PRY SPRY
zfp6	3	Turkey	B-locus zinc finger protein 2	ACA64750	66	0.0	KRAB COG5048
ckr1	2	Green seaturtle	Zinc finger protein CKR1	EMP25061	75	0.0	KRAB COG5048
zfp5	3	Green seaturtle	Zinc finger protein	EMP25063	75	0.0	KRAB COG5048 SFP1
LAO	7	Chicken	L-amino-acid oxidase	BAJ53002	89	0.0	Amino_oxidase

TRIM7.1	7	Turkey	TRIM protein 7	ACA64755	94	0.0	RING BBOX BBC PRY SPRY
Hep21	3	Chicken	Hep21 protein precursor	NP_989852	85	9e-64	none found
TRIM39.2	6	Chicken	TRIM protein 39	BAJ53005	87	0.0	RING BBOX PRY SPRY
TRIM27.2	7	Turkey	TRIM 27	ACA64758	79	0.0	RING BBOX COG4467 PRY SPRY
TRIM39.1	6	Chicken	TRIM protein 39	BAJ53007	77	1e-141	PRY SPRY
TRIM27.1	6	Chicken	TRIM 27	AAW82327	83	0.0	RING BBOX PRY SPRY
TRIM-like	8	Mallard	Zinc finger protein RFP	EOB02164	71	0.0	RING BBOX PRY SPRY
TRIM41	6	Turkey	TRIM protein 41	ACA64761	94	0.0	RING BBOX BBC PRY SPRY
GNB2L1	8	Human	Guanine nucleotide binding protein β-2-like1	EAW53696	100	0.0	WD40
BTN1	44	Turkey	B-butyrophilin 1	ACA64763	73	0.0	PRY SPRY
BTN2	9	Turkey	B-butyrophilin 2	ACA64764	70	2e-115	BBC SPR SPRY KRAB

zfp4	2	Mallard	Zinc finger protein 271	EOA93444	92	0.0	COG5048
zfp3	3	Mallard	Zinc finger protein 208	EOA93445	88	0.0	COG5048
BG2	43	No significant homology					IG_MOG_like V-set COG5048
zfp2	4	Mallard	Zinc finger and SCAN domain-containing protein 2	EOA94931	89	0.0	COG5048
BG1	2	Green seaturtle	Butyrophilin subfamily 2 member A1	EMP39643	73	3e-78	IG_MOG_like V-set
BTN1-like	7	Chicken	Butyrophilin subfamily 1 member A1	NP_001029989	62	4e-53	PRY SPRY
zfp1	2	Mallard	Zinc finger and SCAN domain-containing protein 2	EOA94844	59	2e-111	COG5048
COL11A2	Partial [d]	Mouse	Collagen alpha-2(XI) chain precursor	Q64739	85	0.0	Collagen COLFI
DAA	4	Mallard	MHC class II alpha chain	AEH96282	87	5e-143	MHC_II_alpha IgC

DAB	6	Swan goose	MHC class II beta chain	ACH81993	85	2e-147	MHC_II_beta IgC
DBA1	4	Mallard	MHC class II alpha chain	ADM53418	81	5e-131	MHC_II_alpha IgC
DBB1	6	Swan goose	MHC class II beta chain	ACH81993	83	7e-142	MHC_II_beta IgC
DBA2	4	Mallard	MHC class II alpha chain	ADM53418	81	5e-131	MHC_II_alpha IgC
DBB2	6	Thin-billed prion	MHC class II beta chain	ACO36253	86	8e-138	MHC_II_beta IgC
DBA3	4	Mallard	MHC class II alpha chain	ADM53418	81	5e-131	MHC_II_alpha IgC
DBB3	6	Thin-billed prion	MHC class II beta chain	ACO36253	86	1e-137	MHC_II_beta IgC
BRD2	Partial [d]	Japanese quail	Serine threonine Kinase	BAC82511	89	0.0	Bromo
DMA	Partial [d]	Mallard	MHC class II M alpha chain	EOA93964	80	2e-99	MHC_II_alpha IgC
DMB1	5	Turkey	MHC class II M beta chain 1	ACA64776	70	5e-109	IgC

DMB2	Partial [d]	Greater prairie chicken	MHC class II M beta chain 2	AGC96181	79	1e-118	IgC
UAA	7	Red-billed gull	MHC class I antigen	ADW65600	87	1e-179	MHC_I IgC
TAP1	12	Golden pheasant	Transporter associated with antigen processing 1	AFN53621	79	0.0	ABC_membrane ABC_ATPase
TAP2	9	Turkey	Transporter associated with antigen processing 2	ACA64780	75	0.0	ABC_membrane ABC_ATPase
LTB4R1	Partial [d]	Chicken	Leukotriene B4 receptor 1	BAF63011	81	2e-125	7tm_1
TNXB	Partial [d]	Mallard	Tenascin-X	EOA94117	76	0.0	FN3
UBA	Partial [d]	Red-billed gull	MHC class I antigen	ADW65600	84	2e-171	MHC_I IgC

UCA1	7	Red-billed gull	MHC class I antigen	ADW65600	84	2e-166	MHC_I IgC
ADPRH 1	6	Rock pigeon	[Prtein ADP-ribosylarginine] hydrolase	EMC80157	80	3e-171	ADP_ribosyl_GH
ADPRH 2	8	Rock pigeon	[Prtein ADP-ribosylarginine] hydrolase	EMC80157	83	0.0	ADP_ribosyl_GH
UCA2	7	Red-billed gull	MHC class I antigen	ADW65600	84	9e-168	MHC_I IgC
ADPRH 3	4	Rock pigeon	[Prtein ADP-ribosylarginine] hydrolase	EMC80157	84	4e-99	ADP_ribosyl_GH
UDA	7	Red-billed gull	MHC class I antigen	ADW65600	82	9e-166	MHC_I IgC
OR	1	Green seaturtle	Olfactory receptor 11A1	EMP42676	76	1e-135	7tm_4 7tm_1
PHD finger protein7-like	11	Chicken	PHD finger protein 7-like	XP_425998	67	3e-114	zf-HC5HC2H

USP42-like [e]	1	No significant homology

[a] Excludes two *TAP1* pseudogenes.

[b] Species in which the highest scoring matches were found.

[c] Expect value.

[d] Partial CDSs (coding sequences) were obtained due to gaps.

[e] Predicted CDS was similar to the mRNA sequence of ubiquitin specific peptidase 42 (*USP42*) by BLASTN; however, no significant homology was determined by BLASTP.

Table S3. Long and accurate PCR primers used to verify assembly of the complex MHC region showing tandem repeated "αβ" units

Fragment	Primer name	Primer location	Primer sequence (5′→3′)	Size (bp)	Extension time	Excluded segment
COL11A2-DAA	P1-F	*COL11A2*	CCTGCTTCCCAGTATCACGTTCCTGC	4071	5 min	*COL11A2–*
	P1-R[1]	*DAA, DBA1/2/3* (exon 4)	ATCATCGTGGGCACCATCCTCATCATCA			*DBA1/2/3*
DAA-DAB	P2-F	*DAA* (3'UTR)	TGTGGGCTGGGTACTGGGACCAGTGCAG	3979	5 min	*DAA–*
	P2-R[1]	*DAB, DBB1/2/3* (intron 2)	TGCAAGGAGGTCAGGGAAGGGCGCAGGGGC			*DBB1/2/3*
DAB-DBA1-DBB1	P3-F	*DAB* (exon 2)	AACAGCGGTTTTCCAGGAGCTGGCTGT	5437	7 min	*DAB–...*
	P3-R[1]	*DAB, DBB1/2/3* (exon 2)	GCCCCACATCGCTGTCGAAGTGC			*–DAB*
DBA1-DBB1-DBA2/	P4-F	*DBA1/2/3* (intron 2)	CCTGTTGTGCCTCACTGCCCACCCGGG	4963	6.5 min	*DBA1/2–...*
DBA2-DBB2-DBA3	P4-R[1]	*DAA, DBA1/2/3* (exon 3)	CTGCCCACCGCCCTCCTGAAGCACT			*–DAA*
DBB1-DBA2-DBB2	P5-F	*DBB1* (exon 2）	GAGCGGGTGAGGTATGTGAAGAGGTA	5473	7 min	
	P5-R	*DBB2/3* (exon 2)	CCACCGCAGCCCGTTTCTGCTCCAGTAT			

DBB2-DBA3-DBB3	P6-F	*DAB, DBB1/2* (3'UTR)	AGCTCAGCTTTGTTGTCACTGCTCTGCA	5224	7 min
	P6-R	*DBB3* (3'UTR)	GCAGGGTTGGCAGCAGCGGGTAGGAC		
DBB3-BRD2	P7-F	*DAB, DBB1/2/3* (exon 2)	CCCCCCTGGGCGAGCCCCAAGCCAAGT	4800	6.5 min
	P7-R	*BRD2*	TCACCCGCGGCCTCCCGTGGCTGTCACCTA		

[1]The reverse primers P1–P4 were IIα- or IIβ-universal since we hoped to prove the absence of some segments simultaneously (no corresponding PCR product), and these putatively inexistent segment are listed in the column “Excluded segment.”

Table S4. Primers for quantitative real-time PCR (qRT-PCR) amplification and full-length cDNA sequence isolation

Application	Locus	Location	Primer name	Primer sequence (5′→3′)	Size (bp)	*T*a (℃)
qRT-PCR	*UAA*	Exon 1	A-F	GCCGTCGTCGGGGCGGCG	229	58
		Exon 2	A-R	TGCTCTGTGCGATCTGGGTCTG		
	UBA	Exon 5	B-F	TTGCCGTCATTGCTGGATTCGCCT	153	58
		3’UTR	B-R	TTGCTTTGCACTAGCAGTTCTCTAC		
	UCA1/2	Exon 2	C-F	ACACCGAGACCCGGAACTTTC	171	58
		Exon 3	C-R	CGTCGTAGCCCATCTGCC		
	UDA	Exon 6	D-F	GTACGGCGTGGCGTCAGGGATC	118	58
		3’UTR	D-R	AGCCCCCACCGTAGCTGGGACTG		
5'RACE	*DAB*	Exon 3	IIB1-5'R	AAGGTCTCCCCCATGCAGTCGAGCTCC	470	64
	DBB1/2/3	Exon 3	IIB2-5'R	CTCGACTGCACGGGGTAGATTTC	463	60
	DAA	Exon 3	IIA1-5'R	CCCCCAGCTCCACAGGGTCCTCAGA	451	60
	DBA1/2/3	Exon 3	IIA2-5'R	CCCCCAGCTCCACTTGGCGTTTGGG	451	62
	UAA/UBA	Exon 3	UABA-5'R	TGGTGCTACCGTCATAAAGTGTG	432, 438	58
	UCA1/2	Exon 3	UCA-5'R	CATCTGCCAATATCCCGTGTGTGT	457	62

	UDA	Exon 3	UDA-5'R	GCATCCTGCCGCCCTCCCGTGTGT	462	61
3'RACE	*DAB*	Exon 2	IIB1-3'F	GTTTTCCAGGAGCTGGCTGTGT	889, 934	62
	DBB1/2/3	Exon 2	IIB2-3'F	TACGTACAGTTCCAGTTTAAGTGC	889	61
	DAA	Exon 2	IIA1-3'F	CAACGCGATAATCCAGACTGACTTGT	1266	64
	DBA1/2/3	Exon 2	IIA2-3'F	GCACACGATCCACCAAGCTGAGTTCC	1008	63
	UAA	Exon 2	UAA-3'F	GGGACACCCAGACCCAGATCG	1116	62
	UBA	Exon 2	UBA-3'F	CAGGTTGCTCTTGCGGAAAGAGCCCACACA	1379	59
	UCA1/2	Exon 2	UCA-3'F	ACCCGGAACTTTCTGGGACCACACA	1119	60
	UDA	Exon 2	UDA-3'F	ACTGGGTGAGAGGGAGCCCTCACACA	1037	61

Figure S1. Genomic location (a) and electrophoresis (b) of seven long and accurate PCR primers (P1–P7). Only the shortest fragments were obtained for forward and/or reverse primers with non-specific binding. P4 amplified two different fragments with the same length due to the high sequence similarity among *DBA*s *1–3*. M: DL2504 DNA marker (NormalRunTM 250bp-IV DNA ladder, Generay Biotechnology).

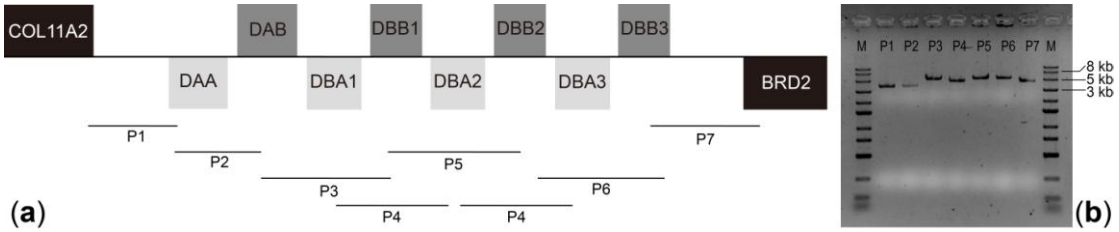
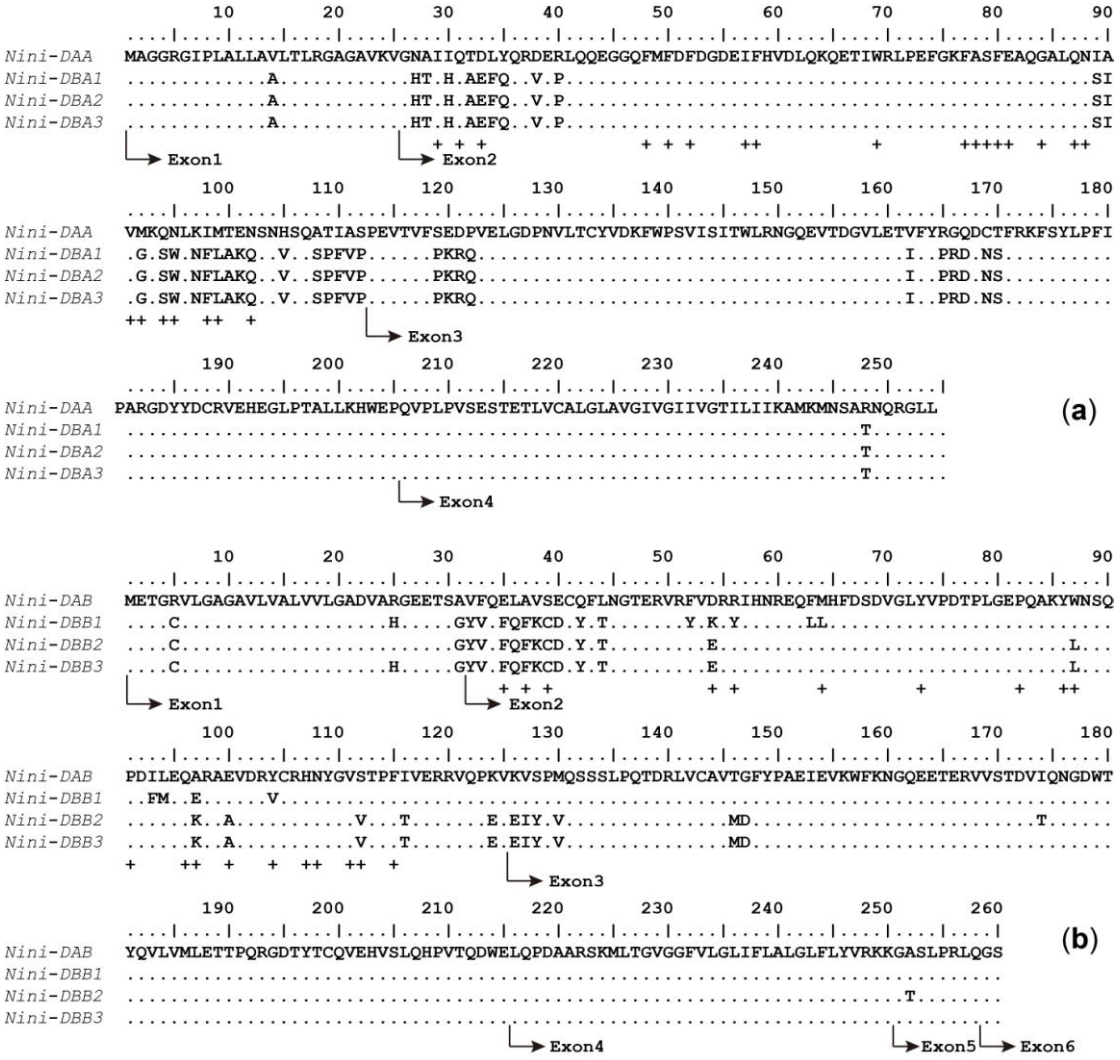


Figure S2. Full-length amino acid alignments of IIα (a) and IIβ genes (b). Dots indicate identity to the first sequence. Crosses represent putative antigen-binding sites inferred from Brown *et al.* [S2] and Stern *et al.* [S3].



Supplementary References

[S1] Ye, Q., He, K., Wu, S.-Y. & Wan, Q.-H. Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. *PLoS One* **7**, e32154, doi:10.1371/journal.pone.0032154 (2012).

[S2] Brown, J. H. *et al*. Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39 (1993).

[S3] Stern, L. J. *et al*. Crystal-structure of the human class-II MHC protein HLA-DR1 complexed with an influenza-virus peptide. *Nature* **368**, 215-221 (1994).