The Journal of Immunology

RESEARCH ARTICLE | JUNE 01 2007

Extended Gene Map Reveals Tripartite Motif, C-Type Lectin, and Ig Superfamily Type Genes within a Subregion of the Chicken *MHC-B* Affecting Infectious Disease [12] FREE

Takashi Shiina; ... et. al

J Immunol (2007) 178 (11): 7162–7172.
https://doi.org/10.4049/jimmunol.178.11.7162

Extended Gene Map Reveals Tripartite Motif, C-Type Lectin, and Ig Superfamily Type Genes within a Subregion of the Chicken *MHC-B* Affecting Infectious Disease[1,2]

Takashi Shiina,* W. Elwood Briles,† Ronald M. Goto,‡ Kazuyoshi Hosomichi,* Kazuyo Yanagiya,* Sayoko Shimizu,* Hidetoshi Inoko,* and Marcia M. Miller[3‡]

MHC haplotypes have a remarkable influence on whether tumors form following infection of chickens with oncogenic Marek's disease herpesvirus. Although resistance to tumor formation has been mapped to a subregion of the chicken *MHC-B* region, the gene or genes responsible have not been identified. A full gene map of the subregion has been lacking. We have expanded the *MHC-B* region gene map beyond the 92-kb core previously reported for another haplotype revealing the presence of 46 genes within 242 kb in the Red Jungle Fowl haplotype. Even though *MHC-B* is structured differently, many of the newly revealed genes are related to loci typical of the MHC in other species. Other *MHC-B* loci are homologs of genes found within MHC paralogous regions (regions thought to be derived from ancient duplications of a primordial immune defense complex where genes have undergone differential silencing over evolutionary time) on other chromosomes. Still others are similar to genes that define the NK complex in mammals. Many of the newly mapped genes display allelic variability and fall within the *MHC-B* subregion previously shown to affect the formation of Marek's disease tumors and hence are candidates for genes conferring resistance. ***The Journal of Immunology,*** **2007, 178: 7162–7172.**

The MHC has long been recognized as an immune-regulatory system important in regard to disease susceptibility and resistance. Autoimmune disease associations with *HLA* are sometimes reported for alleles at individual loci, but with few exceptions it is difficult to understand whether individual loci or other genes nearby are actually responsible for the observed effect. In the chicken, *MHC-B* haplotypes, until recently distinguished primarily by serological methods, display strong influences in several infectious diseases (1–5). Resistance to Marek's disease tumors was mapped to a gene, or genes, closely linked to serologically-defined *MHC-B* class I (as opposed to BG Ags) in analyses of r*MHC-B* haplotypes (6–8). Because *MHC-B* haplotypes reproducibly influence the incidence of Marek's disease in experimental lines and these haplotypes are subject to periodic recombination, it may be possible to determine the contribution of individual genes to disease incidence and make progress in defining pathogen-host interactions that drive MHC gene evolution.

As in many other species, the MHC gene complex in the chicken is characterized by the presence of class I, II, and III genes. However, the organization of these genes in the chicken is different from that in other organisms studied so far. Unlike the large single complexes found in humans and many other mammals or the highly dispersed arrangement of MHC genes found in fish (9), the MHC in the chicken is divided into two major regions, *MHC-Y* (*Y*) and *MHC-B* (*B*), located on the same microchromosome (GGA16) but separated by a crossover breakpoint that results in *Y* and *B* haplotypes assorting independently at meiosis (10). In the *Y* region, distinctive, well-expressed class I genes are intertwined with specialized class II and C-type lectin-like loci (Refs. 11–13, M. M. Miller, R. Goto, T. Shiina, and H. Inoko, unpublished data). In the *B* region, a small minimal core of 66 kb is made up of class I and II genes found in linkage with additional genes typically resident within the MHC including two genes typical of a class III region (14). What lies beyond the class I, II, and III genes found at the core of the *B* region is only partially known. There are two C-type lectin-like genes linked with the core (14, 15). One is a C-type lectin-like locus (*Blec2* or *BNK*) that is well-expressed in NK cells and is structurally similar to human NKR-P1. The second lectin locus (*Blec1*) is likely an early activation Ag. Not yet fully sequenced but known to be tightly linked with *B* class I and II by means of serology is the large family of highly polymorphic Ig superfamily (IgSF)[4] genes encoding BG Ags (sometimes called zipper proteins) that are well-expressed on erythrocytes and other tissues (16–19). BG ectodomains share sequence similarities with

*Division of Basic Medical Science and Molecular Medicine, Department of Molecular Life Science, Tokai University School of Medicine, Kanagawa, Japan; †Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115; and ‡Division of Molecular Biology, Beckman Research Institute, City of Hope National Medical Center, Duarte, CA 91010

Received for publication November 21, 2006. Accepted for publication March 19, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by National Science Foundation MCB-9604589, National Cancer Institute R21 CA105426, National Research Initiative Grant Nos. 2004-35205-14203 and 2006-35205-16678 from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service, and Kakenhi (Grant-in-Aid for Scientific Research) on Priority Areas "Comparative Genomics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan. BAC filters were provided through the U.S. Department of Agricultural Cooperative State Research, Education, and Extension Service Multi-State Research Committee NRSP-8.

[2] The sequence(s) presented in this article has been submitted to GenBank under accession number(s) AB268588.

[3] Address correspondence and reprint requests to Dr. Marcia M. Miller, Beckman Research Institute, City of Hope National Medical Center, 1450 East Duarte Road, Duarte, CA 91010. E-mail address: mamiller@coh.org

[4] Abbreviations used in this paper: IgSF, Ig superfamily; BTN, butyrophilin; MOG, myelin oligodendrocyte glycoprotein; TRIM, tripartite motif; RJF, Red Jungle Fowl; BAC, bacterial artificial chromosome; CDS, coding sequence; BLAST, basic local alignment search tool; EST, expressed sequence tag; LTR, long terminal repeat; SNP, single nucleotide polymorphism; WGS, Whole Genome Shotgun.

mammalian MHC butyrophilin (*BTN*) and myelin oligodendrocyte glycoprotein (*MOG*) (20). A short distance away (5.6 cM) from the minimal core is a partially defined region containing a class IIα gene locus (21). Tripartite motif (*TRIM*)-like genes, two *CD1* genes, and a tenascin-like gene (*TNXB*, also called *TN-Y*) were mapped recently to the *B* vicinity by regional sequencing and physical mapping, but a full map of *B* is still lacking (22–25).

With the goal of defining the genes responsible for the strong association between *B* haplotype and resistance to infectious disease, we set out to expand the *B* gene map and determine which genes lie within the subregion affecting Marek's disease tumors. We obtained the sequence for an interval spanning the region from *BG3* (representing the BG Ag family) to *CD1A1* within the Red Jungle Fowl (*RJF*) haplotype and analyzed this in the context of MHC gene maps for other species. To evaluate interhaplotype variability in gene structure and polymorphism, we compared the *RJF* haplotype to the previously published sequence from the *B12* haplotype. Using the map as a guide, we localized the crossover breakpoint identifying genes within the region affecting Marek's disease tumors thereby defining a subgroup of genes that may be under selection by this viral pathogen.

Materials and Methods

Construction of a bacterial artificial chromosome (BAC) based contig map

Membrane-filter sets for four BAC libraries were provided by J. Dodgson (U.S. Department of Agriculture-Cooperative State Research, Education, and Extension Service, National Animal Genome Research Program Poultry Genome Project, Michigan State University, East Lansing, MI). The libraries (TAM31, TAM32, TAM33, and CHORI-261) were all made with DNA from the same highly inbred UCD001 line RJF hen used in the Chicken Genome Whole Genome Shotgun (WGS) Project. The MHC-*B* haplotype in this animal is referred to here as the *RJF* haplotype. Filters were screened and positive clones were selected following the recommended procedures. Probes included *178/179f* (AF493429), a *B* class I gene-specific PCR product; *bg11* (AF493427); a *chCD1-2* cDNA clone (AY375530) provided by C. Dascher (Brigham and Women's Hospital, Boston, MA); and cloned PCR fragments for MHC class II (DQ007238) and for the complement *C4* locus (DQ007237). Positive clones were end-sequenced, fingerprinted, and analyzed in Southern hybridizations.

Long-range PCR amplifications of CD1 *region*

Two *CD1* PCR products (PCR-CD1A1 and PCR-CD1A2) were obtained using genomic DNA from the line UCD001 RJF hen as template. *CD1A1*- and *CD1A2*-specific primers were designed based on AY849318, CD1A1F (5′-GGGAATATGCAAGGTGATTAACAGCG-3′) and CD1A1R (5′-GGATGTACTTTGACCCACACGCAGC-3′), and CD1A2F (5′-TTGAG GTGAATGGAGCGTGGAATAAA-3′) and CD1A2R (5′-CTGACTGC CGTGGCCCTTGGTT-3′). PCR amplifications were performed with a TaKaRa LA-*Taq* kit (TaKaRa) following the protocol recommended by the manufacturer.

DNA sequencing and analysis

Three BAC clones and two long-PCR products that covered 242 kb from the *BG3* to *CD1A1* genes were completely and bidirectionally shotgun sequenced with an average redundancy of 7.5, which was sufficient for assembly and analysis of the entire sequence using previously established procedures (26, 27). The completed sequence was compared with previously published chicken genomic sequences from the *B12* haplotype found in GenBank accession nos.: AL023516, AY849318, and AY694127 (14, 22, 23). Sequence alignments were performed and homologies were determined using the programs contained within the GENETYX version 11 software packages (www.sdc.co.jp/genetyx). These analyses were complemented with basic local alignment search tool (BLAST) and SwissProt searches for homology and conserved domain sequences, GENSCAN for the prediction of coding sequences, and Repeatmasker2 (http://repeatmasker.genome.washington.edu/) for the identification and classification of repeat sequences.

Diversity analysis

The nucleotide diversity profile was constructed after determining the percent nucleotide difference between the *RJF* and *B12* haplotype sequences for a sliding window of 1 kb with 100-bp overlaps. The diversity profile was then drawn using the graphics output of Microsoft Excel. All indels were removed from the alignments to standardize the number of nucleotides examined within each window. Nonsynonymous to synonymous substitution rates (dN/dS) were calculated by the Nei and Gojobori method (28) with the P-distance parameter in the MEGA3.1 software (www.megasoftware.net).

Haplotype breakpoint mapping

BR5 resulted from a crossover between *B21* and *B19* providing a new haplotype detected serologically as *BG19* and *BF21* (6). To verify the genotypes of *BR5* DNA samples, they were analyzed in the context of parental haplotypes (*B19* and *B21*) for *BG* by Southern hybridization (29), and for MHC I and MHC II by single-strand conformational polymorphism patterns (30). Genotypes for LEI0258 were defined using a LEI0258 primer pair (31). For single nucleotide polymorphism (SNP) genotyping, primer pairs were designed to amplify segments of 500–700 bp at varying intervals between the primary markers (*BG*, *LEI0258*, and *BL*). Informative markers defining the crossover breakpoint in *BR5* were revealed with the following two sets of primer pairs: F11: AGCTCTACCCCAAGCCTG and R11: TGAGAAAACCCCCGTGAGG; and F12: TTGTTGTGGCTCT CAGCCTTC and R12: AAACCGTCCTTCAACCATTCC. All animals providing DNA for this study were maintained under a protocol approved by the Northern Illinois University Institutional Animal Care Review Committee.

Results

Many MHC-related genes reside nearby the MHC-B *core*

Sequence covering 241,833 bp of *B* was obtained from three overlapping RJF BAC clones and two PCR products (Fig. 1*A*). The *RJF* sequence encompasses newly sequenced regions on GGA16 lying on either side of the 92-kb sequence previously reported for *B12*, i.e., as described in Fig. 1*B*, 88 kb upstream and 39 kb downstream, as well as the full sequence of the overlapping RJF region. Interspersed repeats (15 long interspersed elements/CR1 elements and 4 long terminal repeat (LTR) retrotransposons) (Fig. 1*C*) and 37 simple repeated nucleotide motifs (Fig. 1*D*) are distributed across the sequence. Overall, the sequence is extraordinarily G+C-rich (average 55.4%) (Fig. 1*E*), suggesting that GGA16 may be one of the very most G+C-rich among microchromosomes that as a group have a broad distribution of G+C content shifted to higher values than macrochromosomes (32). Within the 242 kb, some segments are even more G+C-rich, such as 33 kb from *KIFC1-LOC417038* (59.6%) and the 69 kb from *BLB1-CYP21* (61.9%). Other segments are relatively more AT-rich, e.g., the 17-kb *BG3-BG2* interval (G+C content 45.6%), as well as the 10 kb that is marked by the LEI0258 microsatellite and contains *B-BTN1* (G+C content 40.4%). Portions of this 242-kb sequence were used to guide assembly of the *B* contigs in the first and second assemblies of RJF (*Gallus gallus*) WGS sequence. *B* remains partially assembled in the WGS sequence and is represented by 18 contigs with 92 segments in the second assembly.

GENSCAN and BLAST searches predict the presence of 46 genes densely packed within the 242-kb sequence (averaging 5.4 kb/gene) (Fig. 1*B* and Table I). The *RJF* haplotype is fully concordant in gene number and spacing with *B12* sequence in the overlapping portions (AL023516, AY849318, and AY694127). The 46 genes include 32 "expressed genes" as assessed by their sequence matching cDNA or expressed sequence tag (EST) clones. Based on structural integrity, an additional nine genes are considered to be "candidate genes." Four more encode tRNA. Donor/acceptor sites are conserved at all intron/exon boundaries in all 32 expressed genes and nine candidate genes except for part of *C4* and *TNXB* where donor/acceptor sites are lacking (data not

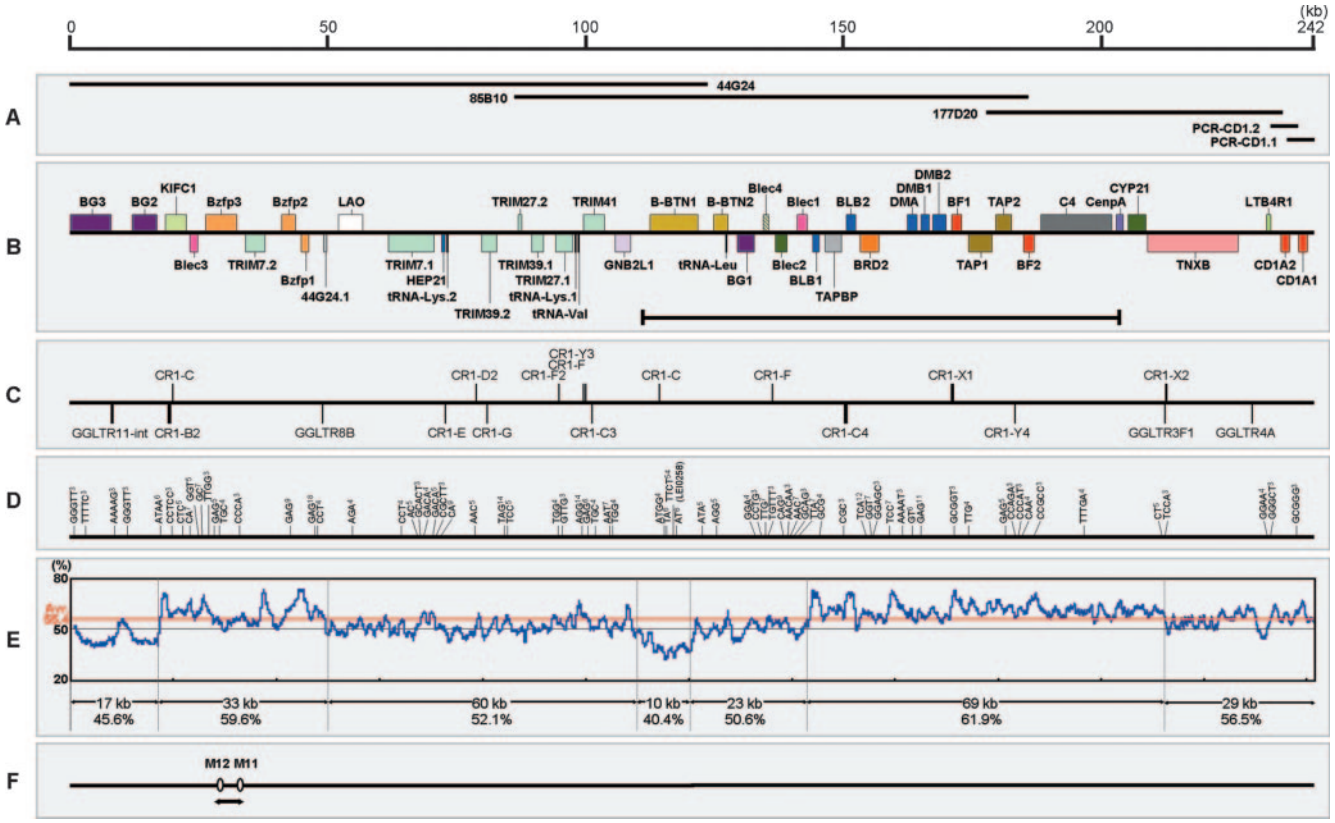

FIGURE 1. Structure of the complete 242 kb (243,833 bp) *RJF* MHC-*B* cluster from *BG3* to *CD1A1*. *A*, The sequence-ready contig constructed from the overlapping set of three BACs (44G24, 85B10, and 177D20) and two PCR products (PCR-CD1A2 and PCR-CD1A1). *B*, Gene content. Genes in the same family are indicated by similarly colored boxes. Upper boxes indicate genes oriented from *BG3* side to *CD1A1* (from *left* to *right* in this figure). Lower boxes indicate genes oriented from *CD1A1* to *BG3*. Bar with blocked ends denotes gene region previously sequenced in *B12*. *C*, The positions of repeat elements, such as L1 and LTR families, are represented by black bars. *D*, Locations of di-, tri-, tetra- and penta-nucleotide repeats are defined individually. *E*, Plot of local GC nucleotide content in overlapping 200 bp windows. *F*, Margins of the *BR5* recombination breakpoint zone are defined by SNPs M11 and M12.

shown). One pseudogene, *Blec4*, found in both *RJF* and *B12* haplotypes (in AL023516 without annotation), is identified by a single exon present within a 1.94-kb inverted repeat that is 96% identical with a segment of the adjacent *Blec2* (*BNK*) locus (Table I). At least half the genes can be considered as contributing to immunity. A total of 18 genes are newly added on the *B* gene map.

Thirteen new genes (*tRNA-Lys-2*, *HEP21*, *TRIM7.1*, *LAO*, *44G24.1*, *Bzfp1*, *Bzfp2*, *TRIM7.2*, *Bzfp3*, *Blec3*, *KIFC1*, *BG2*, and *BG3*) (Fig. 1*B* and Table I) lie upstream of the previously described *TRIM39.2* locus (22). All new genes, except for one, share significant similarities with genes in other species (Table II). Only *44G24.1* is difficult to relate to other genes. *44G24.1* is expressed, but we found no significant homology between it and other genes in database searches. The new *tRNA-Lys-2* locus is the fourth tRNA gene to be mapped. Two new TRIM genes bring the number of TRIM genes to seven (*TRIM41*, *TRIM27.1*, *TRIM39.1*, *TRIM27.2*, *TRIM39.2*, *TRIM7.1*, and *TRIM7.2*). All encode 30.2 domains (33), as does *B-BTN1*, one of the two *BTN* loci previously placed in the *B* extended region. The amino acid similarities between the *TRIM*-like genes and homologous sequences in nonavian species ranged from 55 to 70% (Table II).

There are three zinc finger protein genes (*Bzfp1*, *Bzfp2*, and *Bzfp3*) in the new sequence. Two, *Bzfp1* and *Bzfp2*, share sequence similarity with human genes (Table II). *Bzfp2*, which has a significant similarity to a gene on human 19p13.1-p12, is likely expressed. *Bzfp1* is also likely expressed and is nearly identical (99% amino acid similarity) to the chicken Krueppel-type gene *CKR1* (*LOC396247*, NM_205309) (34). *Bzfp1* shares a 59% amino acid similarity with human *ZFP436* (Q9C0F3) located in 1p36, an *HLA* paralogous region (35). *Bzfp3* shares 99% nucleotide identity with a finished chicken EST clone (BX932436) and is most similar to a gene mapping to human chromosome 5.

Other new genes also with similarities to MHC and MHC paralogous genes in other species include *LAO* and *KIFC1* (Fig. 1*B*, Tables I and II). *LAO* is an L-amino acid oxidase precursor sharing a 70% amino acid similarity with a snake venom enzyme (O93364) (36). *LAO* is also similar to a human amino oxidase gene located in the *HLA* paralogous region on human chromosome 19. *KIFC1* has 65 and 61% amino acid similarities with frog C-terminal kinesin 2 (P79955) and human *KIFC1* (Q9BW19), respectively, which are present in the extended class II region in these species.

Blec3 is a newly recognized C-type lectin-like gene lying 113 kb upstream from previously found *Blec1* and *Blec2* (*BNK*) (Fig. 1*B*). *Blec3*, like *Blec1*, shares highly significant amino acid similarity with CD69 mammalian early activation Ags (Table II). *Blec3* also shares highly significant amino acid similarity with a putative C-type lectin protein FPV239 (P14371) encoded within the genome of fowlpox virus.

The position of *BG2* and *BG3* at the upstream margin of the RJF sequence establishes the distance (128 kb) that separates the *BG* gene family from the MHC class I and class II genes with which they are tightly linked (Fig. 1*B*). Although more BAC clones need to be added to the *B* contig and sequenced to define all the *BG* gene

Table I. *Genes identified in the RJF MHC-B*[a]

Location	Gene Symbol	Alias	Orientation	Exons	Gene ID in WGS	Predicted CDS Acc. Num.	Status	Evidence	Immune Defense	Homology or Prominent Features
<1–7747	***BG3***[b]		**(+)**	**>51**	**X**	**X**	**Gene candidate**	**X**	**Likely**	**Similar to BG Ag, no significant homology with other BG seqs**
11888–16560	*BG2*		(+)	39	396417	NM_205436	Exp. gene	U49098	Likely	Similar to BG Ag, 99% nt identity with U49098
17978–22451	*KIFC1*[b]		(+)	14	417037	XM_415326	Exp. gene	12 ESTs		Similar to C-terminal kinesin 2
22871–24523	***Blec3***[b]		**(−)**	**5**	**X**	**X**	**Gene candidate**	**X**	**Likely**	**Similar to C-type lectin-like receptor (CD69 homologous)**
25917–32075	*Bzfp3*[b]	*LOC425771*	(+)	11	425771, 427095	XM_424688	Exp. gene	BX932436		Similar to zinc finger protein, 99% nt identity with BX932436
33439–37664	*TRIM7.2*		(−)	7	425772	XM_423490	Exp. gene	2 ESTs	Likely	Similar to TRIM protein 7
39631–42445	***Bzfp2***[c]	***LOC427726***	**(+)**	**3**	**X**	**X**	**Gene candidate**	**X**		**Similar to zinc finger protein 43**
44163–47701	***Bzfp1***		**(−)**	**3**	**X**	**X**	**Gene candidate**	**X15538**		**Similar to zinc finger protein cKr1, 97% nt identity with X15538**
48704–49471	***44G24.1***[c]	***LOC417038***	**(−)**	**2**	**X**	**X**	**Gene candidate**	**X**		**Unknown**
51748–56754	*LAO*		(+)	7	417039	XM_415327	Exp. gene	5 ESTs		Similar to L-amino acid oxidase
61137–70451	*TRIM7.1*	*TRIM7*	(−)	9	417040	XM_415328	Exp. gene	BX934475	Likely	Similar to TRIM protein 7, 99% nt identity with BX934475
71716–72531	*HEP21*		(−)	3	395192	NM_204521	Exp. gene	AJ416114	Likely	Hep21 protein, 100% nt identity with AJ416114
73652–72725	*tRNA-Lys.2*		(−)	1	X	X	tRNA	X04180		tRNA-Lys, 100% nt identity with X04180
79241–82743	*TRIM39.2*	*TRIM39*	(−)	6	417041	NM_001006196	Exp. gene	AJ720787	Likely	TRIM protein 39, 100% nt identity with AJ720787
84914–87969	***TRIM27.2***	***TRIMX***	**(+)**	**7**	**430359**	**XM_427919**	**Gene candidate**	**X**	**Likely**	**Similar to TRIM protein 27**
89024–91652	***TRIM39.1***	***BR***	**(−)**	**4**	**X**	**X**	**Gene candidate**	**X**	**Likely**	**Similar to TRIM protein 39**
93751–97371	*TRIM27.1*	*TRIM27*	(−)	7	417042	NM_001030671	Exp. gene	BX934244	Likely	TRIM protein 27, 100% nt identity with BX934244
97924–97997	*tRNA-Lys.1*		(−)	1	X	X	tRNA	K00290		tRNA-Lys, 100% nt identity with K00290
98305–98372	*tRNA-Val*		(−)	1	X	X	tRNA	X17513		tRNA-Val, 100% nt identity with X17513
99013–103684	*TRIM41*		(+)	7	417043	NM_001030672	Exp. gene	4 ESTs	Likely	TRIM protein 41
105251–108529	*GNB2L1*	*C12.3*	(−)	8	417044	NM_001004378	Exp. gene	M24193		Guanine nt binding 12.3 protein, 100% nt identity with M24193
112185–121715	*B-BTN1*[b]	*B30.1*	(+)	20	768783	XM_001231905	Exp. gene	4 ESTs	Likely	Similar to butyrophilin 1 and TRIM protein 39
124457–127628	***B-BTN2***[c]	***B30.2***	**(+)**	**5**	**X**	**X**	**Gene candidate**	**X**	**Likely**	**Similar to TRIM protein 27**
127490–127571	*tRNA-Leu*		(−)	1	X	X	tRNA	K00238		tRNA-Leu, 96% nt identity with K00238
129025–133291	*BG1*	*B-G*	(−)	20	417046∼7	XM_415334∼5	Exp. gene	5 ESTs	Likely	BG Ag
135634–136353	*Blec4*		(+)	1	X	X	Pseudogene	X		99% nt identity with Blec2 exon 1 and intron 1
137090–139582	*Blec2*	*NKr*	(−)	6	693259	NM_001044682	Exp. gene	AJ634338	Likely	Lectin-like NK cell surface protein, 99% nt identity with AJ634338
141449–143522	*Blec1*	*Lectin, CLEC2D*	(+)	5	404776	NM_213582	Exp. gene	AJ634334	Likely	C-type lectin-like receptor, 100% nt identity with AJ634334
144346–145709	*BLB1*		(−)	6	724083	NM_001044694	Exp. gene	AY510090	Yes	MHC class II β-chain 1, 96% nt identity with AY510090
146694–150143	*TAPBP*	*Tapasin*	(−)	8	417048	NM_001034816	Exp. gene	AJ005071	Yes	TAP-binding protein (tapasin), 100% nt identity with AJ005071
151180–152531	*BLB2*		(+)	6	693256	NM_001044679	Exp. gene	DQ008585	Yes	MHC class II β-chain 2, 99% nt identity with DQ008585
153883–157644	*BRD2*	*RING3*	(−)	11	417049	NM_001030674	Exp. gene	25 ESTs		Bromodomain-containing protein 2
163273–165359	*DMA*	*B-MA1*	(+)	4	417050	XM_415338	Exp. gene	8 ESTs	Yes	Similar to MHC class II A family, peptide-loading 2
165744–167523	*DMB1*	*B-MB2*	(+)	5	768619	XM_001231665	Exp. gene	3 ESTs	Yes	Similar to MHC class II B family, peptide-loading 1

(*Table continues*)

Table I. (*Continued*)

Location	Gene Symbol	Alias	Orientation	Exons	Gene ID in WGS	Predicted CDS Acc. Num.	Status	Evidence	Immune Defense	Homology or Prominent Features
168127–171063	*DMB2*	*B-MB1*	(+)	5	417051	XM_415339	Exp. gene	BX932520	Yes	Similar to MHC class II B family, peptide-loading 2, 99% nt identity with BX932520
171880–173913	*BF1*		(+)	8	693260	NM_001044683	Exp. gene	AY821520	Yes	MHC class I α-chain 1, 100% nt identity with AY821520
175032–179828	*TAP1*		(−)	11	427727	XM_425302	Exp. gene	AJ843261	Yes	TAP1, 98% nt identity with AJ843261
180384–183420	*TAP2*		(+)	9	427728	XM_425303	Exp. gene	AJ843262	Yes	TAP2, 98% nt identity with AJ843261
185784–187800	*BF2*		(−)	8	768348	XM_001231203	Exp. gene	S78682	Yes	MHC class I α-chain 2, 100% nt identity with S78682
188855–203096	*C4*		(+)	38	429584	NM_001077233	Exp. gene	DQ224386	Yes	Complement factor, C4
203680–205179	*CenpA*	*B-H3*	(+)	4	X	X	Exp. gene	AY205310∼1		Centromere protein A
205680–209526	*CYP21*[b]		(+)	10	429828	XM_427384	Exp. gene	CR406618		steroid 21-hydroxylase, 99% nt identity with CR406618
209629–227297	*TNXB*		(−)	26	396106	XM_001231291	Exp. gene	X99062		Tenascin XB
23169–234263	***LTB4R1***		**(+)**	**1**	**427729**	**XM_425304**	**Gene candidate**	**X**	**Likely**	**Similar to human leukotriene B4 receptor 1 (LTB4-R1)**
235466–237579	*CD1A2*	*CD1.2*	(−)	6	425802	NM_001024582	Exp. gene	AY849320	Yes	CD1A Ag (CD1A)
238837–241251	*CD1A1*	*CD1.1*	(−)	6	425801	XM_423513	Exp. gene	AY849319	Yes	CD1A Ag (CD1A)

[a] Locations were calculated from the BG3 side of the 242-kb contiguous sequence. The assembled loci are classified into three categories of gene status: "expressed gene (Exp. gene)," "gene candidate," and "pseudogene." "Expressed gene" means a gene that is transcribed to mRNA and also has a reliable open reading frame (ORF) and/or a known protein product, with the accession numbers for the mRNA and protein sequences provided. The "gene candidate" is transcribed to mRNA (mRNA sequence accession number provided) but has an unknown or uncertain ORF, and may or may not have an accession number for a protein sequence listed. Regular text in the table (no bold or underlining) indicates genes. Bold values indicate the gene candidate. The underlined values represent the pseudogene.

[b] Five genes—*BG3* (truncated at contig margin), *KIFC1*, *Blec3*, *B-BTN1*, and *CYP21*—were partially predicted by the Genscan program.

[c] Three genes—*B-fp2*, *44G24.1*, and *B-BTN2*—were fully predicted by the Genscan program.

family members, the remaining, tightly linked family members likely lie nearby upstream. In harmony with previously reported *BG* cDNA sequences, *BG2* and *BG3* genes consist of exons encoding IgV-like and transmembrane domains coupled with multiple 21-nt exons encoding a lengthy coiled-coil intracellular region (16). The *BG1* locus found adjacent to *Blec4* (Fig. 1*B*) shares a similar exon organization, but in contrast has two additional distinctive exons at the 3′-end.

HEP21 is a chicken gene which is predominantly expressed in the oviduct producing a protein secreted into hen egg white (37) (Fig. 1*B* and Tables I and II). It shares no significant homology with any known gene. However, *HEP21* is weakly similar to GPI-linked *LY6G5B* which is encoded in the MHC in mammals. Perhaps *HEP21* is one representative of the *LY6* gene family in the chickens.

New genes mapping downstream of the *B* core contribute further to the definition of the chicken class III region (Fig. 1*B*) and confirm the presence of a well-conserved class III region in avian species. The newly obtained sequence revealed the presence of *CYP21* and *TNXB* downstream of *C4* and *CenpA* forms a unit similar to the discrete, sometimes duplicated, genetic unit found in *HLA* (38). The four genes are present in a highly conserved order (*C4*, *CenpA*, *CYP21*, and *TNXB*). *CYP21* shares 57% similarity with the *HLA* cytochrome P450 (P08686), and *TNXB* has 58% similarity with the *HLA* Tenascin-X precursor (P22105). *CenpA,* although missing from *HLA*, is present within the MHC of *Xenopus* species in the same position as *CenpA* in the chicken (39). The remaining gene, *LTB4R1*, revealed in this segment was unexpected, but consistent with the role of class III genes in inflammation. The predicted amino acid sequence for *LTB4R1* matches most closely a human leukotriene B4 receptor (Q15722), encoded on chromosome 14 that mediates chemotaxis (56% amino acid similarity) (40).

The *RJF* map confirms the short distance between *CD1A2* and *BF2* class I genes (48 kb) and that *CD1A2* and *CD1A1* are separated from each other by ∼840 bp (23–25).

Genetic variability stretches beyond the class I and II loci encompassing some loci within the MHC-B extended region

To further investigate the relative stability of the *MHC-B* core and extended region, we explored genomic and allelic variability by comparing the sequences now available for the *RJF* and *B12* haplotypes. We compared a total of 124,489 bp for the *RJF* haplotype (116,217 bp from *TRIM39.1* to *CenpA* and 8,272 bp in the *CD1* region) with the corresponding 123,987 bp previously determined for the *B12* haplotype (115,721 and 8,266 bp) (22, 23, 41). We found differences in gene structure, insertions and deletions (indels), SNPs, and codon substitutions at many loci (Fig. 2). Structural differences were observed in four genes: *B-BTN2*, *TAP1*, *BF1*, and *BG1*; three of these vary by a few nucleotide differences. Namely, in the *RJF* haplotype exon 5 of *B-BTN2* is 3 bp longer; the final exon of *TAP1* is shorter as the result of 14-bp insertion that introduces a stop codon; and exon 1 (signal peptide) in *BF1* is 15 bp longer. A substantial difference of over 400 bp exists between *BG1* alleles with the *RJF* allele containing a near perfect duplication of four exons and associated introns encoding a portion of the coiled-coil region of BG1.

Numerous indels (2227) were also observed (Fig. 2). These are especially frequent in the introns and intergenic regions of *TRIM27.1*, *DMB2*, *TAP2*, and *C4* genes. The average number of indels overall was 11.1 per kb.

Table II. *Comparison of RJF MHC*-B *genes with genes in other species*[a]

Locus	Species[b]	Chromosome Location in Humans[c]	Prominent Features of the Highest Amino Acid Similarity	Amino Acid Acc. Num.	% Positive	*E* Value	Conserved Domain[d]
BG3	Mouse	Extended class I	MOG precursor	Q61885	46	6e-9	TraB_pillus
BG2		**None found**	**No significant homology**				
KIFC1	African clawed frog	Extended class II	C-terminal kinesin 2	P79955	65	3e-140	KISc_C_terminal
Blec3	Mouse	12p13-p12	Early activation Ag CD69	P37217	52	1e-19	CLECT
Bzfp3	Mouse	11q12	Zinc finger protein 91 (Zfp-91)	Q62511	68	4e-54	SFP1
TRIM7.2	Mouse	5q35.3	TRIM protein 7	Q923T7	58	6e-79	RING BBOX PRY SPRY
Bzfp2	Human	19p13.1-p12	Zinc finger protein 43 (Zinc protein HTF6)	P17038	49	4e-67	KRAB Herpes_LMP2
Bzfp1	Human	1p36	Zinc finger protein 436	Q9C0F3	59	4e-89	KRAB COG5048
44G24.1		**None found**	**No significant homology**				
LAO	snake	19q13.3-q13.4	L-amino acid oxidase precursor	O93364	70	2e-149	Amino_oxidase
TRIM7.1	Human	5q35.3	TRIM protein 7	Q9C029	70	7e-139	RING BBOX BBC PRY SPRY
HEP21		**None found**	**No significant homology**				
TRIM39.2	Rat	Classical class I	TRIM protein 39	Q6MFZ5	62	6e-86	RING BBOX PRY SPRY
TRIM27.2	Human	Extended class I	TRIM protein 27	P14373	55	1e-75	RING BBOX PRY SPRY
TRIM39.1	Rat	Classical class I	TRIM protein 39	Q6MFZ5	60	2e-44	PRY SPRY
TRIM27.1	Human	Extended class I	TRIM protein 27	P14373	57	3e-86	RING BBOX PRY SPRY
TRIM41	Human	5q35.3	TRIM protein 41	Q8WV44	69	8e-131	RING BBOX PRY SPRY
GNB2L1	Human	5q35.3	Guanine nucleotide-binding protein β subunit 2-like 1	P63244	100	1e-179	WD40
B-BTN1	Rat	Classical class I	TRIM protein 39	Q6MFZ5	57	2e-39	PRY SPRY
B-BTN2	Human	Extended class I	TRIM protein 27	P14373	58	6e-12	
BG1	Mouse	Extended class I	MOG precursor	Q61885	56	1e-22	IG_like COG3883
Blec2	Human	12p12.3-13.2	Killer cell lectin-like receptor subfamily F, member 1	Q9NZS2	45	3e-09	CLECT
Blec1	Human	12p13-p12	C-type lectin domain family 2 member B	Q92478	64	3e-26	CLECT
BLB1	Human	Classical class II	HLA class II histocompatibility Ag, DRB1-1 β-chain precursor	P04229	70	5e-57	MHC_II_β IGc
TAPBPL	Human	Extended class II	Tapasin precursor	O15533	46	3e-25	IG_like IGc
BLB2	Human	Classical class II	HLA class II histocompatibility Ag, DRB1-8 β-chain precursor	Q30134	67	2e-56	MHC_II_β IGc
RING3	Human	Classical class II	Bromodomain-containing protein 2	P25440	80	0.0	2 BROMO
DMA	Mouse	Classical class II	Class II histocompatibility Ag, M α-chain precursor	P28078	58	1e-37	MHC_II_α IGc
DMB1	Mouse	Classical class II	Class II histocompatibility Ag, M β1 chain precursor	P35737	54	3e-33	IGc
DMB2	Human	Classical class II	HLA class II histocompatibility Ag, DM β-chain precursor	P28068	56	2e-44	IGc
BF1	Mouse	Classical class I	H-2 class I histocompatibility Ag, K-W28 α-chain precursor	P03991	58	2e-62	MHC_I IGc
TAP1	Human	Classical class II	Ag peptide transporter 1	Q03518	67	6e-130	ABC_ATPase
TAP2	Human	Classical class II	Ag peptide transporter 2	Q03519	68	2e-141	ABC_ATPase
BF2	Mouse	Classical class I	H-2 class I histocompatibility Ag, K-K α-chain precursor	P04223	55	1e-57	MHC_I IGc
C4	Rat	Class III	Complement C4 precursor	P08649	54	0.0	A2M_N ANATO A2M C345C
CenpA	Human	2p24-p21	Centromere protein A	P49450	85	1e-16	H3
CYP21	Human	Class III	Cytochrome P450	P08686	57	9e-74	p450
TNXB	Human	Class III	Tenascin-X precursor	P22105	58	0.0	14 FN3 FReD
LTB4R1	Human	14q11.2-q12	Leukotriene B4 receptor 1	Q15722	56	8e-33	7tm_1
CD1A2	Sheep	1q22-q23	T cell surface glycoprotein CD1b-2 precursor	Q29422	53	2e-29	Gc
CD1A1	Human	1q22-q23	T cell surface glycoprotein CD1b precursor	P29016	51	4e-34	MHC_I IGc

[a] Comparisons identifying homologies were made with GENETYX complemented by BLAST and SwissProt searches. Genes without significant match scores are shown in bold.
[b] Species in which highest scoring matches were found.
[c] Location of homologous genes in human genome. Underlined values denote homologs present in the human MHC and its paralogous regions.
[d] Conserved domains revealed in *MHC-B* genes.

SNPs provide a further measure of the genomic variability between haplotypes and between alleles (Fig. 2 and Table III). Overall, there were 1323 nucleotide variations, i.e., 1.08 nucleotide differences per 100 nucleotides (1.08 SNP %) between the *RJF* sequence and that of *B12* in alignments in which the 2227 indels were excluded. The distribution of SNPs is not uniform. Among the 24 genes within overlapping regions of *RJF* and *B12* haplotypes, SNPs are disproportionately distributed to *BG1* (1.89 SNP %) and the polymorphic class I and II Ag-presenting loci *BF1* (3.30 SNP %), *BF2* (3.69%), *BLB1* (2.05%), and *BLB2* (1.92%) (Fig. 2 and Table III). SNPs are also fairly common in *TRIM27.1*, *B-BTN2*, *Blec1*, *DMB2*, *TAP1*, and *TAP2* and relatively less frequent in *TRIM39.1*, *TRIM41*, *DMA*, *CD1A2*, and *CD1A1*. No significant differences were found in the eight remaining genes (Table

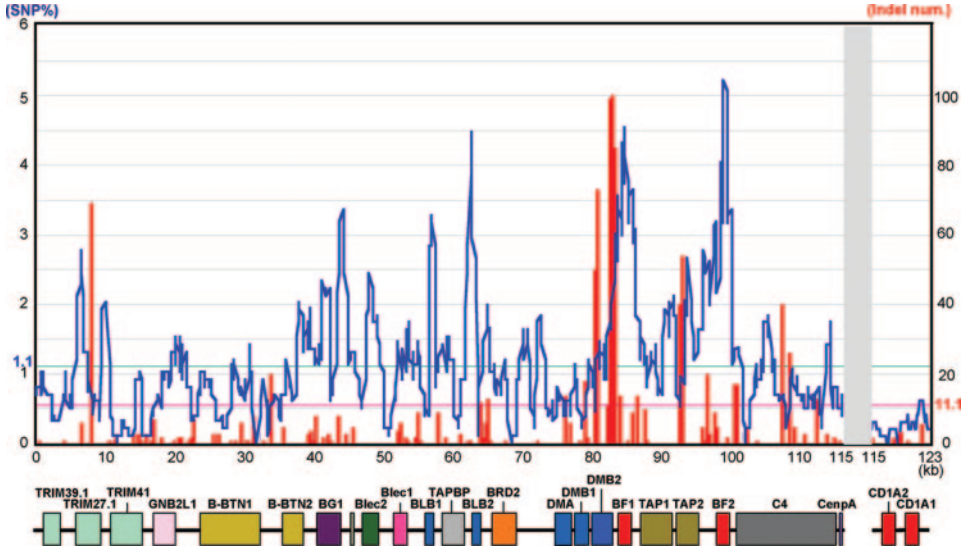

FIGURE 2. Genomic diversity plot of the chicken MHC *RJF* and *B12* haplotypes. Diversity plot drawn upon comparison of the total 123-kb segment covering 115 kb between the *TRIM39.1* and a part of *CenpA* genes and the 9-kb region containing *CD1* genes is shown by blue line, and indels between two haplotypes are shown by red sticks.

III). When SNPs were examined with respect to synonymous and nonsynonymous nucleotide substitutions within coding regions, four genes (*B-BTN2*, *BLB1*, *BLB2*, and *DMB1*) stood out as more frequently containing amino acid-altering substitutions (dN/dS ratios >1). In the same comparison, the dN/dS ratios for the class I loci *BF1* and *BF2* were under 1 (0.9064 in *BF1* and 0.7015 in *BF2*) (Table III). However, when exons 2 and 3 (encoding the BF Ag-binding region) of these loci were examined the dN/dS ratios were significantly higher, 1.4385 for *BF1* and 1.8966 for *BF2*, indicating a large number of synonymous substitutions elsewhere in these loci contribute to their overall low dN/dS ratios.

These data provide ample evidence for genetic variability within the classical Ag processing and presentation loci (*BLB1*, *BLB2*, *DMB1*, *DMB2*, *BF1*, *TAP1*, *TAP2*, and *BF2*) of the *B* core. Significantly, the data also reveal allelic polymorphism within the genes of the extended region. Most prominent are *BG1*, *TRIM27.1*, *B-BTN-2*, and *Blec1* differences. Genetic variability in these and the other genes within the extended region that remain to be analyzed may be contributing to *B* disease associations presently linked to *B* at the level of a haplotype.

*Meiotic crossover breakpoints map to the MHC-*B *extended region*

With the extended map assembled, it was a matter of interest to understand which of the newly mapped genes, if any, map into the *B* region previously shown to influence the incidence of Marek's

Table III. *SNP analysis of 24 genes within the chicken MHC-*B *region where RJF and B12 haplotypes overlap*[a]

Gene	Nucleotide Length in *B12*	Nucleotide length in RJF	Num. Indel	Aligned Sequence Length	Total SNP	Total SNP%	SNP in Intron	SNP in CDS	Synonymous SNP	Nonsynonymous SNP	dS	dN	dN/dS
TRIM39.1	2,629	2,629	0	2,629	14	0.53	10	4	3	1	0.0134	0.0018	0.1343
TRIM27.1	3,683	3,621	76	3,614	44	**1.22**	33	11	8	3	0.0186	0.0028	0.1505
TRIM41	4,677	4,672	11	3,669	15	0.41	13	2	2	0	0	0.0039	*
GNB2L1	3,280	3,279	21	3,269	31	0.95	29	2	2	0	0.0073	0	0
B-BTN1	9,656	9,537	497	9,348	64	0.70	61	3	1	2	0.0046	0.0028	0.6087
B-BTN2	3,175	3,172	3	3,172	43	**1.35**	33	10	6	4	0.0028	0.0057	**2.0357**
BG1	3,773	4,267	512	3,764	71	**1.89**	51	20	9	11	0.0326	0.0143	0.4387
Blec2	2,493	2,493	0	2,493	25	1.00	10	15	0	15	0	0.0288	*
Blec1	2,071	2,074	5	2,070	25	**1.21**	20	5	4	1	0.0256	0.0025	0.09766
BLB1	1,364	1,364	0	1,364	28	**2.05**	3	25	6	19	0.0137	0.0381	**2.7810**
TAPBPL	3,458	3,450	14	3,447	34	0.99	20	14	7	7	0.0172	0.0079	0.4593
BLB2	1,365	1,352	13	1,352	26	**1.92**	5	21	4	17	0.0097	0.0326	**3.3608**
BRD2	3,764	3,762	2	3,762	22	0.58	6	16	14	2	0.0219	0.0013	0.05926
DMA	2,073	2,087	14	2,073	11	0.53	8	3	3	0	0.0122	0	0
DMB1	1,826	1,820	6	1,820	11	0.60	6	5	2	3	0.0044	0.0055	**1.2500**
DMB2	2,900	2,937	143	2,847	32	**1.12**	25	7	5	2	0.0314	0.0123	0.3917
BF1	2,032	2,034	21	1,998	66	**3.30**	30	36	11	25	0.0545	0.0494	0.9064
TAP1	4,808	4,797	12	4,769	63	**1.30**	35	28	21	7	0.0385	0.0059	0.1532
TAP2	3,075	3,037	118	2,997	43	**1.43**	18	25	14	11	0.0208	0.0077	0.3702
BF2	2,013	2,008	5	2,008	74	**3.69**	20	54	20	34	0.0459	0.0322	0.7015
C4	14,251	14,242	131	14,181	121	0.85	90	31	17	14	0.0118	0.0039	0.3305
CenpA	613	612	3	611	5	0.82	3	2	2	0	0	0	0
CD1A2	2,116	2,114	2	2,114	2	0.09	2	0	0	0	0	0	0
CD1A1	2,411	2,415	6	2,410	9	0.37	6	3	1	2	0.0030	0.0026	0.8667
Total	83,506	83,775	1,615	81,781	879	1.12	537	342	161	181	-	-	-

[a] We calculated the number of synonymous or nonsynonymous SNPs of exonic regions using the mRNA and predicted coding sequences included in Table II. "Indel" shows the number of inserted and deleted nucleotides between the sequences. "SNP%" corresponds to the total number of substitutions divided by the length of the aligned sequence and multiplied by 100. Bold values indicate the genes having significantly high SNP% ($p < 0.05$) and high dN/dS ratios. *, Undefined.

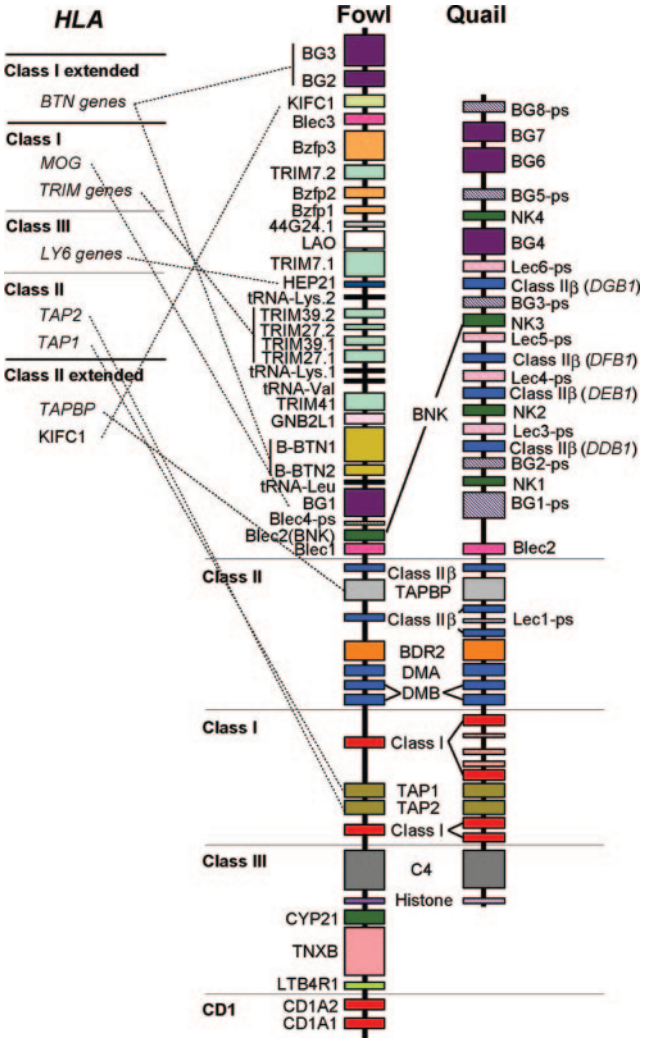

FIGURE 3. Gene maps for human (*HLA*), chicken (MHC-*B*), and quail MHC illustrate the pliant nature of MHC gene organization. Shared presence of *TRIM*, *BTN*, *MOG*, and *LY6* gene family members in the MHC is evident albeit with considerable differences in locale between *HLA* and chicken. Conserved gene content and order is found in the class I-III gene regions of chicken and quail MHC (genes named between maps). The conserved class I-III regions contrast with the remaining portions of the upstream sequences where similarities in gene order are not evident and genes cannot be paired with certainty (genes named on either side of maps). Among the genes upstream only chicken *Blec2 (BNK)* and quail *NK3* are paired based on sequence similarity. Quail MHC region sequence was previously reported by Shiina et al. (56). Pseudogenes are indicated by striped boxes and by *ps* notation. Gene size and spacing is approximate and not to scale.

disease. We analyzed the *BR5* recombinant haplotypes with which resistance to Marek's disease was first mapped to the class I (*BF*) subregion (6). After first verifying the parent of origin for the *BR5* alleles at *BG*, *BL*, and *BF* loci, as well as microsatellite LEI0258 using established methods (see *Materials and Methods*), we used SNPs revealed by limited resequencing to assign the origin of different portions of the *BR5* to one or the other parental haplotype. The *BR5* breakpoint, defined by SNPs M11 and M12, is located upstream of *TRIM7.2* (Fig. 1*F*). Hence, all the genes intervening between the *BR5* breakpoint and the class I genes encoding the BF Ags map within the region affecting Marek's disease. Within this region are a number of genes that could contribute to resistance to the challenge of viral infection beyond the loci for Ag-presenting molecules including the *BG1*, *TRIM27.1*, *B-BTN-2*, and *Blec1* loci that display significant polymorphism as described above.

*Distinctive features of the MHC-*B *gene map support models for MHC evolution in which different regions restructure and evolve at different rates*

Comparisons of *B* with *HLA* and the quail MHC (Fig. 3) reinforce previous findings that a number of different types of genes, in addition to those for Ag presentation and the well-conserved class III region genes, remain in the MHC in divergent species over evolutionary time (9). Comparison of *B* with *HLA* (Fig. 3) shows that many genes characteristic of the different subregions of *HLA* are also present in *B* but reside in the newly mapped region extending upstream of the *B* core. These include: *KIFC1* found in the class II extended segment of *HLA*; *HEP21* the possible homolog of the class III region *Ly6* genes; and the *BTN*, *TRIM*, and *BG* which are all structurally related to genes found with *HLA* class I and class I extended regions. Like *HLA*, *B* also contains zinc finger protein genes even though these appear to be more closely related to genes on other human chromosomes, as is also true some of the *B* TRIM genes. It appears that perhaps over evolutionary time many genes have migrated from a region in which they were commingled, as represented by the *B* extended region, to various discrete locations nearby class I, II, and III genes as found in *HLA* or to other chromosomes.

Comparison of the *B* map with the current map for the quail MHC suggests that the MHC extended regions in birds evolve more rapidly than the core region containing class I, II, and III genes. In Fig. 3, the maintenance of gene order and identity between chicken and Japanese quail is readily apparent across the core class I, II, and III regions even though the genes are unlikely to be orthologous and there are small variations in gene number. In contrast, there are striking differences between RJF and quail maps in the extended region. Aside from shared similarity between the two C-type lectin-like (RJF *Blec1* and quail *Blec2*) genes near the class II region boundary (see Fig. 3), there is little evidence for conservation of gene order and type in the extended regions. Although additional sequencing may reveal in quail genes equivalent to those in the *RJF* extended region beyond the current map, many of the *B* extended region genes found in the *RJF* haplotype are entirely missing in quail in the region directly adjacent to the class I, II, and III core. In contrast to the *BG1*, *GNB2L1*, *BTN*, *TRIM*, *LAO*, zinger finger, *HEP21*, and *KIFC1* genes found in the *RJF* map, the quail map contains instead a series of *BLβ*, *Lec*, and *BNK* genes and pseudogenes that are perhaps derived by repeated duplication within the subregion amid a series of *BG* genes and pseudogenes. It could be that this portion of the MHC in avian species is rapidly restructured as the result of pathogen selection.

Discussion

The striking influence of *B* in Marek's disease has long served as the outstanding example for the expected, but difficult to find, relationship between the MHC polymorphism and infectious disease (42). Previous studies focused on the *B* core region, containing class I, II, and II genes, postulate and effectively reasoned on the basis of the data then available that genes within the *B* core are responsible for the MHC disease associations so evident in chickens (14, 43, 44). The sequencing of the extended *B* region described here has revealed the presence of many more genes in the *B* region in linkage with the previously described core. It is important therefore to understand whether these genes contribute to the remarkable influence of *B* haplotypes in infectious disease. Because many of the newly revealed chicken genes are genes typically found within the MHC region in other species, our findings

show that the *B* region is indeed larger than previously deemed. The extended region sequencing reveals that in the chicken, even more than in quail, the core is compact and minimal, in part, as the result of the avian versions of typically intertwined MHC-associated genes in chickens residing instead within an adjacent region. A number of these MHC-associated genes contribute to immunity, but for others there is no obvious function in immunity (such as the zinc finger protein genes). Understanding which *B* genes are under selection by the challenge of pathogens is important for understanding how the MHC evolves. This work provides a context for additional experiments aimed at this goal. The position of the crossover breakpoint in the *BR5* recombinant haplotype defines which *B* loci need to be considered as possibly involved in the responses to Marek's disease. At this time, it is possible only to suggest which among the mapped genes are likely candidates based on their variability and relatedness to genes of known function in other organisms. Studies of additional recombinant haplotypes will likely narrow the range even possibly allowing the contributions from individual loci to be determined.

TRIM and *BTN* loci are interesting candidates to consider. Both *TRIM* and *BTN* genes encode B30.2 domains. In old world monkeys, the B30.2 domain of TRIM 5α contains a major determinant restricting HIV-1 infection (45). A number of additional TRIM family members display innate immune antiviral properties (46, 47). How many additional TRIM genes also contribute to immunity is not known because the TRIM gene family is large and diverse. Many TRIM genes remain to be studied individually. So it is not yet clear whether any of the seven TRIM genes within *B* have similar properties. They belong to TRIM gene family clusters that have yet to be found to affect viral infection. And, with respect to Marek's disease, *TRIM* genes have not been shown to influence herpesvirus infections. At the same time, one of the three *B TRIM* genes in the *RJF* and *B12* overlapping region, *TRIM27.1*, shows significant SNP variability and nonsynonymous coding region differences (Table III). So variability is present in at least one of the seven *B* TRIM loci. A similar indirect argument can be made for the gene candidate *B-BTN-2*. *B-BTN-2* appears to encode an IgSF cell surface receptor similar to BTNL2, but bearing also a cytosolic B30.2 domain. Although the function of BTN molecules remains to be defined overall, BTNL2 have be shown recently to modulate intestinal inflammation (48) and so it is perhaps more likely that BTN molecules also affect immune fitness in some manner. *B-BTN-2* show high SNP percentage values and a significant dN/dS value suggesting it is under selection. More sequence data from other alleles in different haplotypes will help to reveal how much variability exists at these loci and the other TRIM (*TRIM27.2*, *TRIM39.2*, *TRIM7.1*, and *TRIM7.2*) loci that remain to be compared between haplotypes.

Other candidate genes include the C-type lectin loci. *Blec2* (*B-NK*) is likely a receptor expressed on chicken NK cells and therefore *Blec2* (*B-NK*) is a good candidate for affecting early responses to Marek's disease virus. Our data showing 15 SNPS in the *Blec2 (B-NK)* (Table III) all resulting in amino acid differences between *B12* and *RJF* is consistent with this locus being under selection. Interestingly, this finding corroborates an earlier observation by Rogers et al. (49) showing that all 13 nucleotide changes found in the sequences of *Blec2(B-NK)* in seven different haplotypes produced only nonsynonymous changes. Furthermore, *Blec1*, a potential cell activation marker, also shows significant SNP variability (Table III), but because only a single SNP is nonsynonymous it remains to be seen how polymorphic this locus may be. Blec3 must wait for evaluation because only the single allele sequence is available.

The final candidate locus to consider within extended region is *BG1*. *BG1* is a locus peculiar to chickens that shows substantial sequence differences among alleles including *RJF*, *B12*, and alleles in other haplotypes (R. M. Goto and M. M. Miller, unpublished data). This cell surface protein displays an unusual pattern of variability. Variability is largely confined to duplications and deletions in different haplotypes of four small coiled-coil domains. Because *BG1* is only weakly similar to genes in mammals, it is difficult to infer much from this perspective about possible function; however, this IgSF locus shows distinctive differences among haplotypes indicating that this locus should be included among candidate loci affecting disease incidence.

Five loci within the *B* core region, *DMB2*, *BF1*, *TAP1*, *TAP2*, and *BF2*, show high SNP values (1.12~3.69) in comparisons between the *RJF* and *B12* haplotype (Table II). Therefore, these genes may be affected by the positive selection that generates MHC polymorphisms and associated disease resistance or susceptibility as has been suggested for the dominantly expressed *BF2* class I locus (44). We observed that almost all SNPS in the coding sequence (CDS) of *BF2* occur in the Ag-binding region encoding exons 2 and 3 (45 of 54 SNPs). Interestingly, for the second, minor class I locus *BF1*, almost all the CDS SNPs are also within exons 2 and 3 (25 of 36 SNPs in *BF1*) suggesting that this locus may also be affected by positive selection. We observe similar distributions of CDS SNPs to exons for the Ag-presenting regions in the two class II β loci, 20 of 25 SNPs in *BLB1*, and all SNPs in *BLB2*, suggesting that alleles at these loci may also be positively selected.

Although it is possible perhaps that diversity at the *DMB2*, *TAP1*, and *TAP2* also derive by positive selection, it may be that variability at these loci is not the result of direct selection, but is rather the result of their location in close proximity with the highly selected, polymorphic *BF* genes (Fig. 2) (50). In *HLA*, four locations near the *HLA-A*, *-B*, *-DR/DQ*, and *-DP* loci are thought to be regions affected by hitchhiking and these segments confer susceptibility to several autoimmune diseases such as insulin-dependent diabetes mellitus, rheumatoid arthritis, and psoriasis vulgaris (50, 51). Similar indirect selection might result in the observed variability of *DMB2*, *TAP1*, and *TAP2* whether or not the variability influences disease susceptibility.

The presence in *B* of C-type lectin-like loci, sharing homology with mammalian lymphoid inhibitory (*Blec2-BNK*) and activating (*Blec1* and *Blec3*) receptors strengthens the argument that the ancestral MHC may have contained several classes of receptors and ligands used in the primitive immune defense. Overall, the organization of MHC genes in the chicken and the quail is consistent with recent models of the MHC evolving first as a protocomplex containing genes involved in innate immunity (35, 52) and with the rapid evolution of gene-encoding receptors and ligands for innate and adaptive immune cell interactions. Furthermore, the similarities between the chicken *B* (and *Y*) C-type lectin-like genes and sequences present within the genomes of avian pathogens are striking (M. M. Miller, unpublished data). Implied in these similarities is the use of pirated *B* and *Y* gene sequences as decoys for immune evasion. If this has indeed occurred, then it is not surprising that C-type lectin-like genes evolve rapidly across species and evolutionary time.

The comparison of *B* with the MHC of the Japanese quail suggests that the MHC maybe often be restructured in birds (Fig. 3). Even if *TRIM*, *BTN*, zinc finger genes, and other loci are found later with further sequencing to be located at a more distant point in the quail MHC, it is apparent that the organization of MHC genes in these two closely similar species has already diverged. Ongoing literature suggests that the number, and likely the linkage relationship, of class I and II genes in birds is variable (53–55).

The comparisons presented here are among the first to reveal the composition of the MHC-extended region of the MHC in birds and to suggest that this region is rapidly remodeled. These findings help to strengthen the need to consider the function of these MHC-associated genes in immunity.

A number of MHC genes assigned to chicken chromosome 16 remain to be mapped. Still to be done is completion of the link between *B* and *Y* and finding the location of additional MHC genes, such as the class IIα gene and more genes expected in the class III region. The sequence of the highly polymorphic *BG* gene family awaits completion. It will be interesting to see whether chicken *BG* genes are contiguous or if other genes intertwine as occurs among the *BG* loci in quail.

Acknowledgments

We thank Keely Walker, Renee Kupolos, Linda Yates, Jerzy K. Kulski, Kei Hanzawa, and Jerry Dodgson for many helpful discussions and suggestions for improving data presentation. We thank Jerry Dodgson for generously providing BAC filters and Chris Dascher for providing the *chCD1–2* cDNA clone.

Disclosures

The authors have no financial conflict of interest.

References

1. Schierman, L. W., and W. M. Collins. 1987. Influence of the major histocompatibility complex on tumor regression and immunity in chickens. *Poult. Sci.* 66: 812–818.
2. Taylor, R. L., Jr. 2004. Major histocompatibility (B) complex control of responses against Rous sarcomas. *Poult. Sci.* 83: 638–649.
3. Lamont, S. J., C. Bolin, and N. Cheville. 1987. Genetic resistance to fowl cholera is linked to the major histocompatibility complex. *Immunogenetics* 25: 284–289.
4. Lillehoj, H. S., M. D. Ruff, L. D. Bacon, S. J. Lamont, and T. K. Jeffers. 1989. Genetic control of immunity to *Eimeria tenella*: interaction of MHC genes and non-MHC linked genes influences levels of disease susceptibility in chickens. *Vet. Immunol. Immunopathol.* 20: 135–148.
5. Cotter, P. F., R. L. Taylor, Jr., and H. Abplanalp. 1998. B-complex associated immunity to *Salmonella enteritidis* challenge in congenic chickens. *Poult. Sci.* 77: 1846–1851.
6. Briles, W. E., R. W. Briles, R. E. Taffs, and H. A. Stone. 1983. Resistance to a malignant lymphoma in chickens is mapped to subregion of major histocompatibility (*B*) complex. *Science* 219: 977–979.
7. Plachy, J., V. Jurajda, and V. Benda. 1984. Resistance to Marek's disease is controlled by a gene within the B-F region of the chicken major histocompatibility complex in Rous sarcoma regressor or progressor inbred lines of chickens. *Folia Biol.* 30: 251–258.
8. Hepkema, B. G., J. J. Blankert, G. A. Albers, M. G. Tilanus, E. Egberts, A. J. van der Zijpp, and E. J. Hensen. 1993. Mapping of susceptibility to Marek's disease within the major histocompatibility (B) complex by refined typing of White Leghorn chickens. *Anim. Genet.* 24: 283–287.
9. Kelley, J., L. Walter, and J. Trowsdale. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56: 683–695.
10. Miller, M. M., R. M. Goto, R. L. Taylor, Jr., R. Zoorob, C. Auffray, R. W. Briles, W. E. Briles, and S. E. Bloom. 1996. Assignment of *Rfp-Y* to the chicken major histocompatibility complex/*NOR* microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. *Proc. Natl. Acad. Sci. USA* 93: 3958–3962.
11. Hunt, H. D., R. M. Goto, D. N. Foster, L. D. Bacon, and M. M. Miller. 2006. At least one YMHCI molecule in the chicken is alloimmunogenic and dynamically expressed on spleen cells during development. *Immunogenetics* 58: 297–307.
12. Afanassieff, M., R. M. Goto, J. Ha, M. Sherman, L. Zhong, C. Auffray, F. Coudert, R. Zoorob, and M. M. Miller. 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.
13. Rogers, S., I. Shaw, N. Ross, V. Nair, L. Rothwell, J. Kaufman, and P. Kaiser. 2003. Analysis of part of the chicken *Rfp-Y* region reveals two novel lectin genes, the first complete genomic sequence of a class I α-chain gene, a truncated class II β-chain gene, and a large CR1 repeat. *Immunogenetics* 55: 100–108.
14. Kaufman, J., S. Milne, T. W. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923–925.
15. Rogers, S. L., T. W. Gobel, B. C. Viertlboeck, S. Milne, S. Beck, and J. Kaufman. 2005. Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. *J. Immunol.* 174: 3475–3483.
16. Miller, M. M., R. Goto, S. Young, J. Chirivella, D. Hawke, and C. G. Miyada. 1991. Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex of the chicken. *Proc. Natl. Acad. Sci. USA* 88: 4377–4381.
17. Miller, M. M., R. Goto, S. Young, J. Liu, and J. Hardy. 1990. Antigens similar to major histocompatibility complex B-G are expressed in the intestinal epithelium in the chicken. *Immunogenetics* 32: 45–50.
18. Kaufman, J., and J. Salomonsen. 1993. What in the dickens is with these chickens: an only slightly silly response to the first draft of Langman and Cohn. *Res. Immunol.* 144: 495–502; discussion 502–419.
19. Bikle, D. D., S. Munson, and L. Komuves. 1996. Zipper protein, a B-G protein with the ability to regulate actin/myosin 1 interactions in the intestinal brush border. *J. Biol. Chem.* 271: 9075–9083.
20. Pham-Dinh, D., M. G. Mattei, J. L. Nussbaum, G. Roussel, P. Pontarotti, N. Roeckel, I. H. Mather, K. Artzt, K. F. Lindahl, and A. Dautigny. 1993. Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* 90: 7990–7994.
21. Salomonsen, J., D. Marston, D. Avila, N. Bumstead, B. Johansson, H. Juul-Madsen, G. D. Olesen, P. Riegert, K. Skjodt, O. Vainio, et al. 2003. The properties of the single chicken MHC classical class II α chain (*B-LA*) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. *Immunogenetics* 55: 605–614.
22. Ruby, T., B. Bed'Hom, H. Wittzell, V. Morin, A. Oudin, and R. Zoorob. 2005. Characterisation of a cluster of TRIM-B30.2 genes in the chicken MHC B locus. *Immunogenetics* 57: 116–128.
23. Salomonsen, J., M. R. Sorensen, D. A. Marston, S. L. Rogers, T. Collen, A. van Hateren, A. L. Smith, R. K. Beal, K. Skjodt, and J. Kaufman. 2005. Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proc. Natl. Acad. Sci. USA* 102: 8668–8673.
24. Miller, M. M., C. Wang, E. Parisini, R. D. Coletta, R. M. Goto, S. Y. Lee, D. C. Barral, M. Townes, C. Roura-Mir, H. L. Ford, et al. 2005. Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. *Proc. Natl. Acad. Sci. USA* 102: 8674–8679.
25. Maruoka, T., H. Tanabe, M. Chiba, and M. Kasahara. 2005. Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. *Immunogenetics* 57: 590–600.
26. Shiina, T., G. Tamiya, A. Oka, N. Takishima, T. Yamagata, E. Kikkawa, K. Iwata, M. Tomizawa, N. Okuaki, Y. Kuwano, et al. 1999. Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. *Proc. Natl. Acad. Sci. USA* 96: 13282–13287.
27. Mizuki, N., H. Ando, M. Kimura, S. Ohno, S. Miyata, M. Yamazaki, H. Tashiro, K. Watanabe, A. Ono, S. Taguchi, et al. 1997. Nucleotide sequence analysis of the HLA class I region spanning the 237-kb segment around the HLA-B and -C genes. *Genomics* 42: 55–66.
28. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418–426.
29. Briles, W. E., R. M. Goto, C. Auffray, and M. M. Miller. 1993. A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* 37: 408–414.
30. Goto, R. M., M. Afanassieff, J. Ha, G. M. Iglesias, S. J. Ewald, W. E. Briles, and M. M. Miller. 2002. Single-strand conformation polymorphism (SSCP) assays for major histocompatibility complex *B* genotyping in chickens. *Poult. Sci.* 81: 1832–1841.
31. Fulton, J. E., H. R. Juul-Madsen, C. M. Ashwell, A. M. McCarron, J. A. Arthur, N. P. O'Sullivan, and R. L. Taylor, Jr. 2006. Molecular genotype identification of the *Gallus gallus* major histocompatibility complex. *Immunogenetics* 58: 407–421.
32. Hillier, L. W., W. Miller, E. Birney, W. Warren, R. C. Hardison, C. P. Ponting, P. Bork, D. W. Burt, M. A. Groenen, M. E. Delany, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
33. Rhodes, D. A., B. de Bono, and J. Trowsdale. 2005. Relationship between SPRY and B30.2 protein domains: evolution of a component of immune defence? *Immunology* 116: 411–417.
34. Benn, A., M. Antoine, H. Beug, and J. Niessing. 1991. Primary structure and expression of a chicken cDNA encoding a protein with zinc-finger motifs. *Gene* 106: 207–212.
35. Flajnik, M. F., and M. Kasahara. 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 15: 351–362.
36. Raibekas, A. A., and V. Massey. 1998. Primary structure of the snake venom L-amino acid oxidase shows high homology with the mouse B cell interleukin 4-induced Fig1 protein. *Biochem. Biophys. Res. Commun.* 248: 476–478.
37. Nau, F., C. Guerin-Dubiard, C. Desert, J. Gautron, S. Bouton, J. Gribonval, and S. Lagarrigue. 2003. Cloning and characterization of HEP21, a new member of the uPAR/Ly6 protein superfamily predominantly expressed in hen egg white. *Poult. Sci.* 82: 242–250.
38. Yang, Z., A. R. Mendoza, T. R. Welch, W. B. Zipf, and C. Y. Yu. 1999. Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module): a mechanism for gene deletions and disease associations. *J. Biol. Chem.* 274: 12147–12156.
39. Ohta, Y., W. Goetz, M. Z. Hossain, M. Nonaka, and M. F. Flajnik. 2006. Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *J. Immunol.* 176: 3674–3685.

40. Yokomizo, T., T. Izumi, K. Chang, Y. Takuwa, and T. Shimizu. 1997. A G-protein-coupled receptor for leukotriene B4 that mediates chemotaxis. *Nature* 387: 620–624.
41. Kaufman, J. 1999. Co-evolving genes in MHC haplotypes: the "rule" for non-mammalian vertebrates? *Immunogenetics* 50: 228–236.
42. Klein, J. 1986. *Natural History of the Major Histocompatibility Complex.* Wiley, New York.
43. Kaufman, J., J. Jacob, I. Shaw, B. Walker, S. Milne, S. Beck, and J. Salomonsen. 1999. Gene organisation determines evolution of function in the chicken MHC. *Immunol. Rev.* 167: 101–117.
44. Wallny, H. J., D. Avila, L. G. Hunt, T. J. Powell, P. Riegert, J. Salomonsen, K. Skjodt, O. Vainio, F. Vilbois, M. V. Wiles, and J. Kaufman. 2006. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc. Natl. Acad. Sci. USA* 103: 1434–1439.
45. Stremlau, M., M. Perron, S. Welikala, and J. Sodroski. 2005. Species-specific variation in the B30.2(SPRY) domain of TRIM5α determines the potency of human immunodeficiency virus restriction. *J. Virol.* 79: 3139–3145.
46. Nisole, S., J. P. Stoye, and A. Saib. 2005. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* 3: 799–808.
47. Si, Z., N. Vandegraaff, C. O'Huigin, B. Song, W. Yuan, C. Xu, M. Perron, X. Li, W. A. Marasco, A. Engelman, et al. 2006. Evolution of a cytoplasmic tripartite motif (TRIM) protein in cows that restricts retroviral infection. *Proc. Natl. Acad. Sci. USA* 103: 7454–7459.
48. Arnett, H. A., S. S. Escobar, E. Gonzalez-Suarez, A. L. Budelsky, L. A. Steffen, N. Boiani, M. Zhang, G. Siu, A. W. Brewer, and J. L. Viney. 2007. BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. *J. Immunol.* 178: 1523–1533.
49. Rogers, S., T. Gobel, and J. Kaufman. 2001. Genomic analysis of B-NK and B-lec, two c-type lectin-like genes in the chicken MHC. In the *6th Avian Immunology Research Group Meeting, October 8–10, 2000.* K. A. Schat, ed. American Association of Avian Pathologists, Cornell University, Ithaca, NY, pp. 120–126.
50. Shiina, T., M. Ota, S. Shimizu, Y. Katsuyama, N. Hashimoto, M. Takasu, T. Anzai, J. K. Kulski, E. Kikkawa, T. Naruse, et al. 2006. Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. *Genetics* 173: 1555–1570.
51. Stewart, C. A., R. Horton, R. J. Allcock, J. L. Ashurst, A. M. Atrazhev, P. Coggill, I. Dunham, S. Forbes, K. Halls, J. M. Howson, et al. 2004. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* 14: 1176–1187.
52. Belov, K., J. E. Deakin, A. T. Papenfuss, M. L. Baker, S. D. Melman, H. V. Siddle, N. Gouin, D. L. Goode, T. J. Sargeant, M. D. Robinson, et al. 2006. Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol.* 4: 317–328.
53. Westerdahl, H., H. Wittzell, and T. von Schantz. 2000. MHC diversity in two passerine birds: no evidence for a minimal essential MHC. *Immunogenetics* 52: 92–100.
54. Ekblom, R., M. Grahn, and J. Hoglund. 2003. Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics* 54: 734–741.
55. Bonneaud, C., G. Sorci, V. Morin, H. Westerdahl, R. Zoorob, and H. Wittzell. 2004. Diversity of MHC class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* 55: 855–865.
56. Shiina, T., S. Shimizu, K. Hosomichi, S. Kohara, S. Watanabe, K. Hanzawa, S. Beck, J. K. Kulski, and H. Inoko. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J. Immunol.* 172: 6751–6763.