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

Takashi Shiina, W. Elwood Briles, Ronald M. Goto, Kazuyoshi Hosomichi, Kazuyo Yanagiya, Sayoko Shimizu, Hidetoshi Inoko and Marcia M. Miller

J Immunol 2007; 178:7162-7172; 
doi: 10.4049/jimmunol.178.11.7162
http://www.jimmunol.org/content/178/11/7162

References
This article cites 54 articles, 19 of which you can access for free at:
http://www.jimmunol.org/content/178/11/7162.full#ref-list-1

Why The JI? Submit online.
• Rapid Reviews! 30 days* from submission to initial decision
• No Triage! Every submission reviewed by practicing scientists
• Fast Publication! 4 weeks from acceptance to publication

Subscription
Information about subscribing to The Journal of Immunology is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
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

Takashi Shiina,* W. Elwood Briles,† Ronald M. Goto,‡ Kazuyoshi Hosomichi,* Kazuyo Yanagiya,* Sayoko Shimizu,* Hidetoshi Inoko,* and Marcia M. Miller3‡

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 rMHC-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) 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 Multi-State Research Committee NRSP-8.

2 The sequence(s) presented in this article has been submitted to GenBank under accession number(s) AE028858.

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: mamille@coh.org

www.jimmunol.org

Copyright © 2007 by The American Association of Immunologists, Inc. 0022-1767/07/$2.00
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 IIa 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 CD1AI 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. Hodgson (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 170/179 (AF493429), a 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’-GGGAAAATGCGGGTTTTGATTACCAACGGC-3’) and CD1A1R (5’-GGGATGTACTTTGTACGGGACACGGC-3’), and CD1A2F (5’-TTTGAGGTGATGGCCGAGTGGATGAAT-3’) and CD1A2R (5’-CTGACTGCGTGCGGCCTTTGGTT-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 CD1AI 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).

Haplotypic breakdown mapping

BFR resulted from a crossover between B21 and B19 providing a new haplotype detected serologically as BG19 and BF21 (6). To verify the genotypes of BFR 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 BFR were revealed with the following two sets of primer pairs: F1: AGCTTACCCCAAGCCTG and R1: TGAGAAACCCCGCCTGAGG; and F12: TTGTGTGCTGCTCGACCTTC and R12: AAACCGTCCTTCAACCATTC. All animals providing DNA for this study were maintained under a protocol approved by the Northern Illinois University Institutional Animal Care Review Committee.
amino acid similarity) to the chicken Krueppel-type gene
expressed. Significant similarity to a gene on human 19p13.1-p12, is likely ex-
pressed with human genes (Table II).

Vian species ranged from 55 to 70% (Table II).

between the TRIM genes to seven (33), as does

TRIM39.2, TRIM27.2, Bzfp1

44G24.1

Table I) lie upstream of the previously de-
scribed 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 be-
tween the TRIM-like genes and homologous sequences in nona-
vian 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 signif-
ificant similarity to a gene on human 19p13.1-p12, is likely ex-
pressed. 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.
1B, 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-termi-
nal kinesin 2 (P79955) and human KIFC1 (Q9BW19), respec-
tively, 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. 1B).
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 RIF
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. 1B). Although more BAC clones need

to be added to the B contig and sequenced to define all the BG gene

![FIGURE 1. Structure of the complete 242 kb (243,833 bp) RIF MHC-B cluster from BG3 to CD1A1.](http://www.jimmunol.org/)
<table>
<thead>
<tr>
<th>Location</th>
<th>Gene Symbol</th>
<th>Alias</th>
<th>Orientation</th>
<th>Exons</th>
<th>Gene ID in WGS</th>
<th>Predicted CDS Acc. Num.</th>
<th>Status</th>
<th>Evidence</th>
<th>Immune Defense</th>
<th>Homology or Prominent Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1–7747</td>
<td>BG3b</td>
<td>A</td>
<td>(+)</td>
<td>&gt;51</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to BG Ag, no significant homology with other BG seqs</td>
</tr>
<tr>
<td>11888–16560</td>
<td>BG2</td>
<td></td>
<td>(+)</td>
<td>39</td>
<td>396417 NM_205436</td>
<td>Exp. gene</td>
<td>U49098</td>
<td>Likely</td>
<td>Similar to BG Ag, 99% nt identity with U49098</td>
<td></td>
</tr>
<tr>
<td>17978–22451</td>
<td>KIFC1b</td>
<td>X</td>
<td>(+)</td>
<td>14</td>
<td>417037 XM_415326</td>
<td>Exp. gene</td>
<td>12 ESTs</td>
<td>Likely</td>
<td>Similar to C-terminal kinesin 2</td>
<td></td>
</tr>
<tr>
<td>22871–24823</td>
<td>Blec3</td>
<td></td>
<td>(−)</td>
<td>5</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to C-type lectin-like receptor (C10b homologous)</td>
</tr>
<tr>
<td>25917–32075</td>
<td>Bzfp3b</td>
<td>LOC425771</td>
<td>(+)</td>
<td>11</td>
<td>425771, 427095</td>
<td>Exp. gene</td>
<td>BX932436</td>
<td>Likely</td>
<td>Similar to zinc finger protein, 99% nt identity with BX932436</td>
<td></td>
</tr>
<tr>
<td>33439–37664</td>
<td>TRIM7.2</td>
<td></td>
<td>(−)</td>
<td>7</td>
<td>425772 XM_423490</td>
<td>Exp. gene</td>
<td>2 ESTs</td>
<td>Likely</td>
<td>Similar to TRIM protein 7</td>
<td></td>
</tr>
<tr>
<td>39631–42445</td>
<td>Bzfp2c</td>
<td>LOC427726</td>
<td>(−)</td>
<td>3</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to zinc finger protein 43</td>
</tr>
<tr>
<td>44163–47701</td>
<td>Bzfp1</td>
<td></td>
<td>(−)</td>
<td>3</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to zinc finger protein cKr1, 97% nt identity with X15538</td>
</tr>
<tr>
<td>48704–49471</td>
<td>4G24.1b</td>
<td>LOC417038</td>
<td>(−)</td>
<td>2</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to L-amino acid oxidase</td>
</tr>
<tr>
<td>51748–56754</td>
<td>LAO</td>
<td></td>
<td>(+)</td>
<td>7</td>
<td>417039 XM_415327</td>
<td>Exp. gene</td>
<td>5 ESTs</td>
<td>Likely</td>
<td>Similar to TRIM protein 7, 99% nt identity with BX934475</td>
<td></td>
</tr>
<tr>
<td>61137–70451</td>
<td>TRIM7.1</td>
<td>TRIM7</td>
<td>(−)</td>
<td>9</td>
<td>417040 XM_415328</td>
<td>Exp. gene</td>
<td>BX934475</td>
<td>Likely</td>
<td>Similar to TRIM protein 7</td>
<td></td>
</tr>
<tr>
<td>71716–72531</td>
<td>HEP21</td>
<td></td>
<td>(−)</td>
<td>3</td>
<td>395192 NM_205421</td>
<td>Exp. gene</td>
<td>AJ416114</td>
<td>Likely</td>
<td>Hep21 protein, 100% nt identity with AJ416114</td>
<td></td>
</tr>
<tr>
<td>73652–72725</td>
<td>tRNA-Lys2</td>
<td></td>
<td>(−)</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>rRNA</td>
<td>X04180</td>
<td>Likely</td>
<td>TRIM protein 39, 100% nt identity with AJ720787</td>
</tr>
<tr>
<td>79241–82743</td>
<td>TRIM39.2</td>
<td>TRIM39</td>
<td>(−)</td>
<td>6</td>
<td>417041 NM_001006196</td>
<td>Exp. gene</td>
<td>AJ720787</td>
<td>Likely</td>
<td>TRIM protein 39, 100% nt identity with AJ720787</td>
<td></td>
</tr>
<tr>
<td>84914–87969</td>
<td>TRIM27.2</td>
<td>TRIMX</td>
<td>(+)</td>
<td>7</td>
<td>430359 XM_427919</td>
<td>Exp. gene</td>
<td>5 ESTs</td>
<td>Likely</td>
<td>Similar to TRIM protein 27</td>
<td></td>
</tr>
<tr>
<td>89024–91682</td>
<td>TRIM6.1</td>
<td>BR</td>
<td>(−)</td>
<td>4</td>
<td>470422 NM_001030671</td>
<td>Exp. gene</td>
<td>3 ESTs</td>
<td>Likely</td>
<td>Similar to TRIM protein 39</td>
<td></td>
</tr>
<tr>
<td>93751–97371</td>
<td>TRIM27.1</td>
<td>TRIM27</td>
<td>(−)</td>
<td>7</td>
<td>417042 NM_001003671</td>
<td>Exp. gene</td>
<td>BX934424</td>
<td>Likely</td>
<td>TRIM protein 27, 100% nt identity with BX934424</td>
<td></td>
</tr>
<tr>
<td>97924–97997</td>
<td>tRNA-Lys1</td>
<td></td>
<td>(−)</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>rRNA</td>
<td>K00290</td>
<td>Likely</td>
<td>X20418</td>
</tr>
<tr>
<td>98305–98372</td>
<td>tRNA-Val</td>
<td></td>
<td>(−)</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>rRNA</td>
<td>X17513</td>
<td>Likely</td>
<td>X17513</td>
</tr>
<tr>
<td>99013–103684</td>
<td>TRIM41</td>
<td></td>
<td>(+)</td>
<td>7</td>
<td>470431 NM_001030672</td>
<td>Exp. gene</td>
<td>4 ESTs</td>
<td>Likely</td>
<td>TRIM protein 41</td>
<td></td>
</tr>
<tr>
<td>105251–108529</td>
<td>GNB2LI</td>
<td>C12.3</td>
<td>(−)</td>
<td>8</td>
<td>47044 NM_001004378</td>
<td>Exp. gene</td>
<td>M2193</td>
<td>Likely</td>
<td>Guanine nt binding 13.2, 100% nt identity with M2193</td>
<td></td>
</tr>
<tr>
<td>112185–121715</td>
<td>B-BTN1b</td>
<td>B30.1</td>
<td>(+)</td>
<td>20</td>
<td>768783 NM_001231905</td>
<td>Exp. gene</td>
<td>4 ESTs</td>
<td>Likely</td>
<td>Similar to butyrophilin 1 and TRIM protein 39</td>
<td></td>
</tr>
<tr>
<td>124457–127628</td>
<td>B-BTN2e</td>
<td>B30.2</td>
<td>(+)</td>
<td>5</td>
<td>X</td>
<td>X</td>
<td>Gene candidate</td>
<td>X</td>
<td>Likely</td>
<td>Similar to TRIM protein 27</td>
</tr>
<tr>
<td>127490–127571</td>
<td>tRNA-Leu</td>
<td></td>
<td>(−)</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>rRNA</td>
<td>K00238</td>
<td>Likely</td>
<td>X20418</td>
</tr>
<tr>
<td>129025–133291</td>
<td>BG1</td>
<td>B-G</td>
<td>(−)</td>
<td>20</td>
<td>4704687–7 XM_415334–5</td>
<td>Exp. gene</td>
<td>5 ESTs</td>
<td>Likely</td>
<td>X20418</td>
<td></td>
</tr>
<tr>
<td>135634–136353</td>
<td>Blec4</td>
<td></td>
<td>(−)</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>Pseudogene</td>
<td>X</td>
<td>Likely</td>
<td>X20418</td>
</tr>
<tr>
<td>137090–139582</td>
<td>Blec2</td>
<td>NKr</td>
<td>(−)</td>
<td>6</td>
<td>693259 NM_001044682</td>
<td>Exp. gene</td>
<td>AJ634338</td>
<td>Likely</td>
<td>Lectin-like NK cell surface protein, 99% nt identity with AJ634338</td>
<td></td>
</tr>
<tr>
<td>141449–143522</td>
<td>Blec1</td>
<td>Lectin, CLEC2D</td>
<td>(+)</td>
<td>5</td>
<td>404776 NM_213582</td>
<td>Exp. gene</td>
<td>AJ634334</td>
<td>Likely</td>
<td>C-type lectin-like receptor, 100% nt identity with AJ634334</td>
<td></td>
</tr>
<tr>
<td>144346–145709</td>
<td>BLB1</td>
<td></td>
<td>(−)</td>
<td>6</td>
<td>724083 NM_001044694</td>
<td>Exp. gene</td>
<td>AY510090</td>
<td>Yes</td>
<td>MHC class II β-chain 1, 96% nt identity with AY510090</td>
<td></td>
</tr>
<tr>
<td>146694–150443</td>
<td>TAPBP</td>
<td>Tapasin</td>
<td>(−)</td>
<td>8</td>
<td>470484 NM_001034816</td>
<td>Exp. gene</td>
<td>AJ005071</td>
<td>Yes</td>
<td>TAP-binding protein (tapasin), 100% nt identity with AJ005071</td>
<td></td>
</tr>
<tr>
<td>151180–152531</td>
<td>BLB2</td>
<td></td>
<td>(−)</td>
<td>6</td>
<td>693256 NM_001044679</td>
<td>Exp. gene</td>
<td>DQ008585</td>
<td>Yes</td>
<td>MHC class II β-chain 2, 99% nt identity with DQ008585</td>
<td></td>
</tr>
<tr>
<td>153883–157644</td>
<td>BRD3</td>
<td>RING3</td>
<td>(−)</td>
<td>11</td>
<td>417049 NM_001030674</td>
<td>Exp. gene</td>
<td>25 ESTs</td>
<td>Yes</td>
<td>Bromodomaining protein 2</td>
<td></td>
</tr>
<tr>
<td>163273–165339</td>
<td>DMA</td>
<td>B-MA1</td>
<td>(−)</td>
<td>4</td>
<td>417050 XM_415338</td>
<td>Exp. gene</td>
<td>8 ESTs</td>
<td>Yes</td>
<td>Similar to MHC class II A family, peptide-loading 2</td>
<td></td>
</tr>
<tr>
<td>165744–167523</td>
<td>DMB1</td>
<td>B-MB2</td>
<td>(+)</td>
<td>5</td>
<td>768619 NM_001231665</td>
<td>Exp. gene</td>
<td>3 ESTs</td>
<td>Yes</td>
<td>Similar to MHC class II B family, peptide-loading 1</td>
<td></td>
</tr>
</tbody>
</table>

(Table continues)
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. 1B) 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. 1B 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. 1B) 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.
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 haplotypes and between alleles (Fig. 2 and Table III). Over-
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 disease associations present.
resistance to Marek’s disease was first mapped to the class I (BF) region. The 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.

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. 1F). 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 theBTN, 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 con-mingled, 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.

Comparisons 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 BL, 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 recombiant 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 recombaint 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 5a 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 RIJ 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 BTN2L2, but bearing also a cytosolic B30.2 domain. Although the function of BTN molecules remains to be defined overall, BTN2L2 have been 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 RIJ 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 RIJ, 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, BFI, TAP1, TAP2, and BF2, show high SNP values (1.12–3.69) in comparisons between the RIJ 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 locaons 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 linkage 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 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


