Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions

Takashi Shiina, Sayoko Shimizu, Kazuyoshi Hosomichi, Sakae Kohara, Seiki Watanabe, Kei Hanzawa, Stephan Beck, Jerzy K. Kulski and Hidetoshi Inoko

*J Immunol* 2004; 172:6751-6763; doi: 10.4049/jimmunol.172.11.6751

http://www.jimmunol.org/content/172/11/6751

References This article cites 62 articles, 24 of which you can access for free at: http://www.jimmunol.org/content/172/11/6751.full#ref-list-1

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
Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions\textsuperscript{1,2}

Takashi Shiina,* Sayoko Shimizu,* Kazuyoshi Hosomichi,† Sakae Kohara,*, Seiki Watanabe,† Kei Hanzawa,⩞ Stephan Beck,‡ Jerzy K. Kulski,*§ and Hidetoshi Inoko\textsuperscript{3}*

We mapped two different quail Mhc haplotypes and sequenced one of them (haplotype A) for comparative genomic analysis with a previously sequenced haplotype of the chicken Mhc. The quail haplotype A spans 180 kb of genomic sequence, encoding a total of 41 genes compared with only 19 genes within the 92-kb chicken Mhc. Except for two gene families (B30 and tRNA), both species have the same basic set of gene family members that were previously described in the chicken “minimal essential” Mhc. The two Mhc regions have a similar overall organization but differ markedly in that the quail has an expanded number of duplicated genes with 7 class I, 10 class IIB, 4 NK, 6 lectin, and 8 B-G genes. Comparisons between the quail and chicken Mhc class I and class II gene sequences by phylogenetic analysis showed that they were more closely related within species than between species, suggesting that the quail Mhc genes were duplicated after the separation of these two species from their common ancestor. The proteins encoded by the NK and class I genes are known to interact as ligands and receptors, but unlike in the quail and the chicken, the genes encoding these proteins in mammals are found on different chromosomes. The finding of NK-like genes in the quail Mhc strongly suggests an evolutionary connection between the NK C-type lectin-like superfamily and the Mhc, providing support for future studies on the NK, lectin, class I, and class II interaction in birds. The Journal of Immunology, 2004, 172: 6751–6763.

The genomic region of the MHC (Mhc) consists of a number of multigene family members involved in the immune responses of vertebrates. Class I and class II genes encode glycoproteins that transport foreign peptides to the surface of cells for recognition by TCRs on lymphocytes, which in turn kill infected cells (1). The chicken (Gallus domesticus) Mhc (Gaido-Mhc)\textsuperscript{3} is located within part of the B complex on microchromosome 16 and it is called the B-F/B-L region. This region has been sequenced and found to contain 19 genes within 92 kb of genomic sequence (See Fig. 4; Refs. 2, 3). The encoded genes are thought to have coevolved as haplotypes (3) with a high linkage disequilibrium between the genes due to the compact structure of its “minimal essential” Mhc (4). A large polymorphic family of B-G genes in close proximity to the B-F/B-L region in the chicken may have also evolved with the B-F/B-L region (5, 6). The chicken has another Mhc region, Rfp-Y, on microchromosome 16, possibly as a result of an en bloc duplication and transposition (7–9). This region houses the nonclassical MHC genes but it has not yet been fully sequenced (10, 11). Allograft rejection and susceptibility to the development of Rous sarcoma virus-induced tumors and Marek’s disease lymphoma appear to be controlled by the B-F/B-L region (12–14).

The quail (Coturnix japonica)\textsuperscript{5} belongs to the same order (Galliformes) and family (Phasianidae) as the chicken. Chicken-quail hybrids have been produced by artificial insemination of quails with chicken semen (15). Several immunological traits have been compared between various lines of quail that were selected for high and low IgY levels in the serum (16), and for high and low secondary Ab responses to Newcastle disease virus (17), influenza virus, sheep erythrocytes, and Salmonella pullorum (18). The quail is susceptible to Marek’s disease virus as observed in the chicken. However, the immune response against the Marek’s disease tumor-associated surface Ag in the quail differs significantly from that in the chicken (19). Such immunological differences are likely to be due in part to the number and variation in MHC gene loci and/or alleles. Experiments on xenogenetic transplantation in early embryonic stages between the quail and chicken showed acute rejections when quail tissue was grafted onto chicken, but not when chicken tissue was grafted onto quail (20–22). However, skin grafts performed in either direction in adult animals were rejected according to a similar temporal pattern (22). This difference in immunotolerance in xenotransplantation may be explained by a greater capacity of the Mhc repertoire in the quail than the chicken to become tolerant to Ags presented peripherally from the early developmental stages (22). Our genomic analyses have previously revealed the presence of many more Mhc class I and class IIB genes in the quail genome than in the chicken genome (23, 24). For example, the quail was found to encode at least 12 Mhc class I loci with the gene products for at least four of these loci possibly serving as ligands to TCRs (23, 25, 26). Therefore, the different outcomes of xenogenetic transplantation between chicken and quail

---

\textsuperscript{1}Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Bobseidai, Isehara, Kanagawa, Japan; \textsuperscript{2}Laboratory of Animal Physiology, Faculty of Agriculture, Tokyo University of Agriculture, Atsugi, Kanagawa, Japan; \textsuperscript{3}Laboratory of Human Sequencing/Immunogenomics, Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; \textsuperscript{4}Centre for Bioinformatics and Biological Computing, School of Information Technology, Division of Arts, Murdoch University, Murdoch, Western Australia, Australia.

Received for publication May 30, 2003. Accepted for publication March 15, 2004.

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.

\textsuperscript{5}This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) “Genome Science” from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. S.B. was supported by the Wellcome Trust. S.B. and H.I. gratefully acknowledge receipt of a collaboration grant (S-03168) from the Japanese Society for the Promotion of Science.

\textsuperscript{6}Address correspondence and reprint requests to Dr. Hidetoshi Inoko, Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Bobseidai, Isehara, Kanagawa 259-1193, Japan. E-mail address: inoko@issc.u-tokai.ac.jp

\textsuperscript{7}Abbreviations used in this paper: Gado, Gallus domesticus; Coja, Coturnix japonica; contig, contiguous cosmids; LINE, long interspersed nuclear element; LTR, long terminal repeat.
may be attributed in part to the differences in the number of expressed MHC loci and the associated capacity for Ag presentation.

To better elucidate the molecular structure and gene organization of the quail Mhc and to investigate how the genomic differences between quail and chicken might affect their respective immune responses, we determined the nucleotide sequence of the 180-kb quail Mhc that is possibly in synteny with the 92-kb chicken Mhc (B-F/B-L region).

**Materials and Methods**

**Isolation and assembly of contiguous cosmids (contigs)**

Cosmid clones were isolated by colony-hybridization screening of a previously constructed quail cosmid library (23). Approximately 1.3 × 10⁶ independent colonies derived from this library were individually screened with the Coja class I cDNA clone, QF63, Coja class IIB, and Coja-BG7 PCR products. The Coja class IIB PCR product was obtained with class IIB-specific primers covering the exons 1–3; CIIBA (5′-CCGGTGCTGCCTACGGAAAC-3′) and CIIBB (5′-AACACCTCCTGGAGTCTC-3′), and the Coja-BG7 PCR product was made with BG7-specific primers covering exons 1–3: QBGA (5′-TCAGCCCACACCTCTTCCA-3′) and QBGB (5′-GGCTCTCTCTGAGTCTGCA-3′). PCR primer pairs and probes from PCR products were also made for 29 locus-specific markers based on the end sequences of cosmid clones and for other genes such as RING3, TAPBL, and TAP2. Contigs were assembled by Not I restriction mapping, PCR-based mapping with 29 locus-specific primer pairs, and nucleotide sequence determination by PCR-based sequencing of portions of class I, class IIB, RING3, TAPBL, and TAP2 (Fig. 1).

**Sequencing strategy**

Six cosmids covering a 180-kb segment of the quail Mhc from the Coja-BG7 to Coja-H3 genes were subjected to nucleotide sequence determination by the shotgun method (27). DNA sequencing was performed by the cycle sequencing method using AmpliTaq-DNA polymerase FS, fluorescently labeled BigDye terminators in a GeneAmp PCR system (Applied Biosystems, Foster City, CA). A 377 and 3100 ABI DNA sequencer were used for automated fluorescent sequencing (Applied Biosystems).

**Assembly and genomic analyses**

Individual sequences were minimally edited to remove vector sequences, transferred to a SPARC station (Sun Microsystems, Palo Alto, CA) and assembled into a contig using the GENETYX-Σ/SQ software (Software Development, Tokyo, Japan). Remaining gaps or areas of ambiguity were analyzed by the direct-sequencing procedure using PCR amplification products obtained with appropriate PCR primers. The final sequence was initially analyzed using the GENETYX software (Software Development). The analysis was complemented by using BLAST (www.ncbi.nlm.nih.gov/blast) for homology searches, Genscan (http://genes.mit.edu/GENSCAN.html) for the prediction of coding sequences, Repeatmasker2 (http://repeatmasker.genome.washington.edu/) for the identification and classification of repeat sequences, and Graal for the identification of CpG islands. Prediction of transmembrane regions and primary structures of amino acid sequences of Coja-BG genes were conducted using the SOSUI server (http://sosui.proteome.bio.nat.ac.jp/sosuiframe.html) and Chou-Fasman programs. Dot matrix analyses were performed using the Harrplot 2.0 software (Software Development). Multiple alignment of sequences and construction of phylogenetic trees were conducted using the Clustal W software at DDBJ (http://www.ddbj.nig.ac.jp/).

**RT-PCR cloning and sequence determination**

Total RNA was extracted from quail tissues with the TRIzol-LS reagent (Invitrogen, Groningen, Netherlands) based on the guanidinium thiocyanate method. Coja class IIB gene-specific oligonucleotide primers were used for RT-PCR amplification: the sense primer (5′-CCGGTGCTGCCTACGGAAAC-3′) and the anti-sense primer (5′-AACACCTCCTGGAGTCTC-3′). The unique sequence in the signal peptide region (exon 1) of the Coja class IIB loci was used to design the sense primer and the sequence of the conserved β2 domain region (exon 3) was used to design the antisense primer. RT-PCR was performed by the oligo(dT) primer method following the protocol provided by the manufacturer (TOYOBO, Osaka, Japan). RT-PCR products (10 ng) were amplified by PCR with 2.5 U of Taq NT polymerase (Nippon Gene, Toyama, Japan) using the thermal cycler GeneAmp PCR system 9700 (Applied Biosystems). The reaction mixture (10 µl), containing dNTPs (200 µM), MgCl₂ (2.5 mM), and gelatin (0.01%), was subjected to 35 cycles of 30 s at 96°C, 30 s at 65°C, and 30 s at 72°C. PCR products were cloned into the pCR2.1 vector with the TA cloning kit according to the protocol provided by the manufacturer (Invitrogen) and subjected to nucleotide sequence determination. To avoid

![FIGURE 1](http://example.com/figure1.png) Molecular contig map (contig 1–5) covering 508 kb of the quail Mhc (Coja) region defined by overlapping cosmids clones. Dotted lines and asterisks indicate the locations of locus-specific markers. Lines and circles indicate the locations of class I, class IIB, and the non-MHC genes that were on overlapping clones and subjected to direct sequencing for construction of the contig. Arrows indicate Not I sites.
PCR and sequencing artifacts generated by polymerase errors, ~100 clones per organ were sequenced.

Results
Assembly and sequencing of contigs
To determine the genomic sequence of the quail Mhc region, a sequence-ready contig was constructed by screening a quail cosm id genomic library made from a 3-wk-old male. This male was from an inbred line of high IgY producers (16, 23) and was confirmed in this study to be heterozygous for Mhc alleles. Screening was performed by colony hybridization with the Coja class I cDNA clone, QF63 (25), and PCR products of the Coja class IIB gene, Coja-BG7 gene, and Coja-E 5’ flanking regions as probes. A total of 28 cosm id clones were isolated and assembled into a map of five contiguous sequences (1–5) covering 508 kb by using restriction enzyme Not I mapping, PCR-based mapping with 29 locus-specific markers derived from the end sequences of cosm id clones, and direct sequencing of 29 PCR-based mapping regions. The overlapping clones were also assessed and ordered in part by PCR-sequencing analysis of the class I and class IIB genes, and nonclass I and nonclass II genes such as RING3, TAPBL, and TAP2 (Fig. 1). The nucleotide sequence differences between the class I loci (2–15%), the class II loci (1–28%) and the 29 locus-specific markers (1–8%) within the cosm id clones were used to confirm that the cosm ids were assembled correctly into their respective contig 1–5. Contig 3 and 5 correspond to clusters 1 and 2 that we reported previously (23).

Contig 1 represents the first fully sequenced Coja Mhc haplotype A and it is composed of 10 cosm id clones covering a 180-kb region with 7 class IIB genes (DGB1, DFB1, DEB1, DBB1, DCB1, DBB1, and DAB1), 4 class I genes (D1, D2, B1, and E), TAPBPL, RING3, and TAP2. Contig 2 is part of the other Coja Mhc hap lotype B composed of eight clones covering 130 kb with eight class IIB genes (tentatively designated as IIB9, IIB8, IIB7, IIB6, −IIB5, IIB4, IIB3, and IIB2), and TAPBPL. Contig 3 is also part of the Coja Mhc hap lotype B but it is composed of five clones covering 70 kb with one class IIB gene (tentatively designated as IIB1), six class I genes (D1, D2, B1, B2, E, and C2), RING3, and TAP2 (Fig. 1). The gap between contig 2 and 3 has not been connected due to the absence of overlapping cosm id clones for this region. The size of the gap between the two contigs was estimated to be shorter than 2 kb in length by PCR analysis using one primer set (sense primer: 5'-ACCGCGCACGACGACGAGCCCAGCGCAT-3' and anti-sense primer: 5'-TCTGGAAACCTCTCCTCCTGATCT-3'). The nucleotide sequence of one of the sequence tag markers, CTR in contig 3, showed a perfect match with the LINE sequence identified between the DBB1 and DAB1 genes in contig 1. In addition, the TAPBPL, RING3, and TAP2 genes, which are highly conserved among jawed vertebrates, and the three Not I sites were localized to almost the identical positions in contig 2, suggesting that they are physically linked to each other and are different from that of contig 1. Such different gene organizations between Mhc haplotypes have been well characterized previously in the HLA-DRB gene region of HLA class II haplotypes (28). Contig 4 and 5 are relatively short segments covering 40 kb and 78 kb with two (A and C1) and four (D3, C3, D4, and C4) class I genes, respectively (Fig. 1). The exact locations of contig 4 and 5 on the quail genome, their linkage to each other, and their linkage to other contigs, remains to be established.

Contig 1 (haplotype A), which appears to be in synteny with the chicken Mhc B-F-L region, was selected for genome sequencing (Fig. 2A). Six contigs (C52, C37, C9, C1, SW70, and SW201) were shotgun sequenced to produce 179,985 bp (GenBank accession number BA078884 and Fig. 2B) of continuous sequence linking Coja-BG8 (one of quail orthologous genes of chicken B-G) to histone (quail histone-H3-like gene). This result was obtained with a high redundancy of 7.4. Overlaps between all the cosm id clones were ascertained at the nucleotide level. The G+C content was 54.6%.

FIGURE 2. Gene organization of the 180-kb (179,985 bp) quail Mhc region from the Coja-BG8 to the Coja-H3. A. An operational contig (contig 1) was constructed from a set of six overlapping clones (C52, C37, C9, C1, SW70, and SW201) that were subjected to nucleotide sequencing. B. Gene map. Filled boxes indicate Coja class I and class IIB genes. Open boxes indicate other genes. Arrows indicate the orientations of the genes. C. Location of CpG islands. D, Location of repetitive elements. E, Plot of the local G+C content in overlapping 200-bp windows. A line indicates the average G+C content (54.6%).
54.6% in the quail Mhc region (Fig. 2, C and E) and 56.8% in the chicken Mhc B-F/B-L region (3).

Repeat sequences

Analysis of the complete sequence with the RepeatMasker 2 program detected the following repeats: 8 long interspersed nuclear elements (LINEs), 17 long terminal repeats (LTRs), 2 minisatellite repeats (QMR1 and QMR2), and 37 simple repeats (Fig. 2D). Of the eight LINEs, two were located within the intergenic regions between the DAB1 and Lec1 genes and between the DFB1 and Lec4 genes. Although the Lec2 gene is composed of four exons, Lec1 and Lec4 lack exons 1 and 2, with the LINEs occupying ~3 kb around each of these areas (Fig. 2B and Table I). Two QMR1-LTR and QMR2-LTR sequences were located in tandem within the intergenic regions between DDB1 and BG2, and between DGB1 and BG3, respectively. QMR1 and QMR2 are long imperfect repeat sequences, (ACTGGATTTGCCAGCAGGC)30 and (ATTTATCGATCGTGTGC)4, respectively (Fig. 2D). In contrast, none of these repeat sequences were found in the chicken Mhc B-F/B-L region (3).

Gene identification

The 180-kb genomic sequence stretching from BG8 to histone-H3 was subjected to gene identification analysis using the BLAST, Genescan, and Rummage programs. These analyses revealed 41 genes, corresponding to one gene per every 4.4 kb. This gene density is slightly higher than that in the chicken Mhc B-F/B-L region with one gene per every 4.6 kb (3), but four times higher than in the HLA region with one gene per every 15.6 kb (29). The 41 genes within the quail Mhc include 15 “expressed” genes as assessed by their sequence identity with cDNA or expressed sequence tag clones (six class IIB genes, DGB1, DAB1, DBC1, DGB1, DDB1, and DAB1; four class I genes, D1, D2, B1, and E; and five other genes, TAPBPL, RING3, TAP1, TAP2, and C4). Twelve other genes, based on their structural integrity, were assessed to be “possibly expressed” (four Coja class II genes, DFBI, DMA1, DMB1, and DMB2; three Coja-BG genes, BG7, BG6, and BG4; four Coja-NK genes, NK1 to NK4; and one Coja-Lec gene, Lec2). Fourteen genes were classified as pseudogenes because of the presence of premature termination sequences or lack of exons (three class I genes, F, G, and H, five Coja-BG genes, BG8, BG5, BG3, BG2, and BG1; five Coja-Lec genes, Lec1 and Lec3 to Lec6; and the histone-H3) (Fig. 2B and Table I). Four Coja class I genes, D1, D2, B1, and E have the same isotypic loci in contig 23 or cluster 1 that was previously described (23). All of the other genes are described and designated in this study for the first time. The

Table I. Genes identified in the Coja region

<table>
<thead>
<tr>
<th>Location</th>
<th>Name</th>
<th>Orientation</th>
<th>Exons</th>
<th>Homology or Prominent Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2501–3751</td>
<td>BG8</td>
<td>C</td>
<td>3</td>
<td>Chicken B-G Ag homologous pseudogene</td>
</tr>
<tr>
<td>5575–10530</td>
<td>BG7</td>
<td>C</td>
<td>16</td>
<td>Chicken B-G Ag homologous gene</td>
</tr>
<tr>
<td>11510–16419</td>
<td>BG6</td>
<td>C</td>
<td>15</td>
<td>Chicken B-G Ag homologous gene</td>
</tr>
<tr>
<td>20781–22036</td>
<td>BG5</td>
<td>C</td>
<td>3</td>
<td>Chicken B-G Ag homologous pseudogene</td>
</tr>
<tr>
<td>24559–26666</td>
<td>NK4</td>
<td>C</td>
<td>4</td>
<td>NK cell receptor, B-NK homologous gene</td>
</tr>
<tr>
<td>29019–35257</td>
<td>BG4</td>
<td>C</td>
<td>22</td>
<td>Chicken B-G Ag homologous gene</td>
</tr>
<tr>
<td>37348–38903</td>
<td>Lec6</td>
<td>C</td>
<td>4</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>39418–40627</td>
<td>DGB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>43617–49358</td>
<td>BG3</td>
<td>C</td>
<td>3</td>
<td>Chicken B-G Ag homologous pseudogene</td>
</tr>
<tr>
<td>47271–49400</td>
<td>NK3</td>
<td>C</td>
<td>4</td>
<td>NK cell receptor, B-NK homologous gene</td>
</tr>
<tr>
<td>51316–52874</td>
<td>Lec5</td>
<td>C</td>
<td>4</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>53398–53411</td>
<td>DFB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>56822–57345</td>
<td>Lec4</td>
<td>C</td>
<td>2</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>57865–59073</td>
<td>DEB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>60625–62395</td>
<td>NK2</td>
<td>C</td>
<td>4</td>
<td>NK cell receptor, B-NK homologous gene</td>
</tr>
<tr>
<td>65137–65195</td>
<td>Lec3</td>
<td>C</td>
<td>4</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>66394–67614</td>
<td>DDB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>70673–71928</td>
<td>BG2</td>
<td>C</td>
<td>3</td>
<td>Chicken B-G Ag homologous pseudogene</td>
</tr>
<tr>
<td>74515–76553</td>
<td>NK1</td>
<td>C</td>
<td>4</td>
<td>NK cell receptor, B-NK homologous gene</td>
</tr>
<tr>
<td>78894–85137</td>
<td>BG1</td>
<td>C</td>
<td>22</td>
<td>Chicken B-G Ag homologous pseudogene</td>
</tr>
<tr>
<td>87243–88798</td>
<td>Lec2</td>
<td>C</td>
<td>4</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>89318–90523</td>
<td>DGB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>90168–95885</td>
<td>TAPBPL</td>
<td>C</td>
<td>8</td>
<td>TAP binding protein (Tapasin) like protein gene</td>
</tr>
<tr>
<td>96793–98015</td>
<td>DDB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>98747–99270</td>
<td>Lec1</td>
<td>C</td>
<td>2</td>
<td>C-type lectin, B-lec homologous pseudogene</td>
</tr>
<tr>
<td>101528–102679</td>
<td>DAB1</td>
<td>C</td>
<td>6</td>
<td>MHC class II β-chain gene</td>
</tr>
<tr>
<td>106765–11039</td>
<td>RING3</td>
<td>C</td>
<td>11</td>
<td>Serine threonine kinase gene</td>
</tr>
<tr>
<td>117119–120774</td>
<td>DMA1</td>
<td>C</td>
<td>4</td>
<td>MHC class I A family, peptide-loading gene</td>
</tr>
<tr>
<td>121721–123740</td>
<td>DMB2</td>
<td>C</td>
<td>6</td>
<td>MHC class II B family, peptide-loading gene</td>
</tr>
<tr>
<td>125218–127370</td>
<td>DMB1</td>
<td>C</td>
<td>5</td>
<td>MHC class II B family, peptide-loading gene</td>
</tr>
<tr>
<td>128976–130973</td>
<td>D1</td>
<td>C</td>
<td>8</td>
<td>MHC class I α-chain gene</td>
</tr>
<tr>
<td>131797–132020</td>
<td>F</td>
<td>C</td>
<td>1</td>
<td>MHC class I α-chain pseudogene</td>
</tr>
<tr>
<td>134392–134851</td>
<td>G</td>
<td>C</td>
<td>1</td>
<td>MHC class I α-chain pseudogene</td>
</tr>
<tr>
<td>136420–136552</td>
<td>H</td>
<td>C</td>
<td>1</td>
<td>MHC class I α-chain pseudogene</td>
</tr>
<tr>
<td>140903–142884</td>
<td>D2</td>
<td>C</td>
<td>8</td>
<td>MHC class I α-chain gene</td>
</tr>
<tr>
<td>143875–148092</td>
<td>TAP1</td>
<td>C</td>
<td>11</td>
<td>Transporter associated with Ag-processing gene</td>
</tr>
<tr>
<td>149395–152830</td>
<td>TAP2</td>
<td>C</td>
<td>9</td>
<td>Transporter associated with Ag-processing gene</td>
</tr>
<tr>
<td>153461–155416</td>
<td>B1</td>
<td>C</td>
<td>8</td>
<td>MHC class I α-chain gene</td>
</tr>
<tr>
<td>156634–158601</td>
<td>E</td>
<td>C</td>
<td>8</td>
<td>MHC class I α-chain gene</td>
</tr>
<tr>
<td>159665–176818</td>
<td>C4</td>
<td>C</td>
<td>38</td>
<td>Complement factor gene</td>
</tr>
<tr>
<td>179607–179723</td>
<td>histone</td>
<td>C</td>
<td>1</td>
<td>Histone H3-like pseudogene</td>
</tr>
</tbody>
</table>

* Location is indicated by nucleotide positions numbered from the Coja-BG8 gene. Gene orientation from class IIB to class I region is shown by “+”, whereas “−” depicts the opposite.
donor/acceptor sites for the 15 expressed and 12 possibly expressed genes were conserved in all intron/exon boundaries (data not shown), except for DBB1 (see below). Among the 27 expressed or possibly expressed gene sequences within this region, 26 of them, with the exception of RING3, have a previously determined function in the immune response system.

Coja class I genes

The Coja class I region of haplotype A has seven class I loci with the gene order of D1, F, G, H, D2, B1, and E from the Coja class IIB to class I regions, spanning ~31 kb (Fig. 2B and Table I). The length of the quail class I gene region is twice as long as the length of the chicken class I gene region (16 kb) (3). Gene orientations are the same for D1, F, G, H, and D2, but the reverse for B1 and E. The F, G, and H genes are class I pseudogenes newly identified within the 10-kb segment between the D1 and D2 genes (Fig. 2B and Table I). The gene encompass only 224 bp from intron 6 to exon 8, and has 84.0% nucleotide identity with D2 and E. The G and H genes encompass 460 and 133 bp from exon 1 to intron 1 with 87.9 and 99.2% nucleotide identity to B1, respectively. Thus, several class I pseudogenes are embedded in the quail Mhc region as compared with the absence of Mhc class I pseudogenes in the chicken Mhc (3). Of the four expressed class I genes, D1 and D2 appear to be the nonclassical class I loci in terms of a weak expression and limited tissue-specificity, as described previously for B-FA2 in the chicken Mhc (3).

Phylogenetic trees of the quail and chicken Mhc class I genes showed that the quail class I genes are more closely related to each other than to the chicken class I B-FA1 and B-FA2 and Y-FA genes (Fig. 3A). This suggested that the quail class I genes were generated after speciation of the quail and the chicken from the common ancestor. The locations of D1, F, G, H, and D2 genes in the quail Mhc region correspond to that in the chicken B-FA2 gene region, whereas locations of the quail B1 and ~E genes correspond to the chicken B-FA1 gene region (see Fig. 6B). However, two amino acid insertions observed in the D1 and D2 proteins at amino acid position 53 in the midst of the α1 domain (exon 2) were absent from the quail B1 and E proteins and the chicken FA1 and FA2 proteins. In addition, two amino acid deletions at position 292 of the transmembrane region (exon 5) were observed in the quail D1, D2, and E proteins, but not in the quail B1 or the chicken FA1 and FA2 proteins (25, 26). Therefore, neither D1, D2, nor E are likely to be the orthologous genes of the chicken B-FA1 or B-FA2. In fact, the Coja-B1 displays the highest nucleotide identity (86%) with B-FA1, and they may be the orthologous counterparts.

Coja class IIB genes

The gene order of the 10 Coja class II genes was established within an 88-kb segment from DGB1 to DMB1 (Fig. 2B and Table I). If the class II gene region in the quail Mhc is defined as the genomic area from BG8 to DMB1, then it is 125 kb in length as compared with 42 kb in the corresponding chicken class II gene region (Fig. 4). The gene orientation is the same for DGB1, DBF1, DEB1, DDB1, and DCB1 but reversed for DDB1 and DAB1. All of the class IIB genes, except for DDB1, are composed of six exons with an average length of 1.2 kb, and an extremely compact structure with small intronic sizes (the total length of introns 1–5 is ~400 bp) (Table I). The lengths of the quail genes and introns are similar to those of the chicken class IIB B-LB1 and B-LB2 genes.

RT-PCR analysis of the Coja class IIB loci by cloning and sequencing RT-PCR products revealed DAB1 and DBB1 transcription in four immunological organs, the lymphocytes, bursa of Fabricius, thymus, and spleen. In contrast, transcripts of DCB1, DEB1, DFB1, and DGB1 were found only in one or other of these immunological organs in a tissue-specific manner. No DDB1 transcripts were detected in the four immunological organs that we examined (Table II). Although the transcript of DBB1 was observed in all of the immunological tissues, it contained an additional 30-bp nucleotide sequence corresponding to the hypothetical intron 1 region. The donor site of this hypothetical DDB1 intron 1 region is conserved (GT), but its acceptor site changes from AG to CT, which results in no splicing of the 30-bp genomic fragment. The amino acid sequence deduced from the inserted 30-bp nucleotide sequence does not contain any structural defects such as premature termination and frame shift in codon usage, suggesting that DDB1 consists of five exons (exon 1: signal peptide + additional 10 aa + β1 domain; exon 2: β2 domain; exon 3: transmembrane domain; exons 4 and 5: cytoplasmic domain) that are likely to encode a class IIB-like protein.

Exon 2 encoding the β1 domain is the most variable of the nucleotide and amino acid sequences among the Coja class IIB genes (75.6–97.8% nucleotide identities; 83.3% on average), as similarly observed in the class IIB genes of other species. The 18 putative peptide binding sites, predicted from human and pheasant class IIB structural analyses (30, 31), displayed a high degree of nucleotide variations mostly accompanied by an amino acid exchange, not only among the Coja class IIB loci, but also between the Coja and Gado class IIB loci (Table III). The number of divergent amino acid residues within the peptide binding sites of the Coja class IIB loci ranged between 2 and 15, 11.4 on average, which is higher than the average of eight divergent residues detected between the two Gado class IIB loci, B-LB1 and B-LB2, as listed in Table IV. In contrast, exon 3 that encodes the β2 domain is completely conserved (100% nucleotide identity) and exons 4–6 are well conserved (93.9–100% nucleotide identities) among the Coja class IIB loci, as was similarly observed between the chicken B-LB1 and B-LB2 (data not shown). Interestingly, these well-conserved regions have very high GC contents (~70%) and contain CpG islands (Fig. 2, C and E).

The phylogenetic tree of the quail and chicken class IIB genes (Fig. 3B) showed that the quail genes are more closely related to each other than to the chicken class II genes. This finding suggests that, similarly to the class I genes (Fig. 3A), the quail class IIB genes were also duplicated after species speciation of the quail and the chicken. The nucleotide sequence identities within all of the domains (β1, β2, transmembrane, and cytoplasmic regions) ranged from 90.8 to 99.0% (between the Coja class IIB loci and from 87.9 to 89.8% between the Coja and Gado class IIB loci). The sequence identity was 97.7% between Gado B-LB1 and B-LB2. The DAB1 and DDB1 loci were found within the 11-kb segment between TAPBL and RING3, and this location corresponds to the chicken B-LB2 locus in the Gado Mhc region. Moreover, these two quail genes were expressed in lymphocytes, bursa of Fabricius, thymus, and spleen, similarly to B-LB2 that encodes a major class IIB protein in the chicken. In comparison, the quail DCB1, DDB1, DEB1, and DGB1 genes had limited expression in the immunological organs. These genes are located within a 52-kb segment between Coja-Lec6 and TAPBL and correspond in location to the chicken B-LB1 locus, which is also only moderately expressed (Table II and Ref. 32).

Three nonclassical class II genes, DMA1, DMB1, and DMB2, that possibly encode the class II Ag-processing molecules (33), were identified within the 18-kb segment between the RING3 and D1 genes near the class I and class II boundary region (Fig. 2B). The amino acid sequences of the translated genes DMA1, DMB1, and DMB2 have 82.5, 85.7, and 83.8% similarity with the chicken B-MA1, B-MB1, and B-MB2 molecules, respectively (see Fig. 6D).
The amino acid identity between the Coja-DMB1 and the Coja-DMB2 was considerably lower (52.8%), although six cysteine sites essential for formation of functionally important disulfide bridges were conserved in both molecules. This relatively low sequence identity may indicate that DMB1 and DMB2 have distinct but unknown functions rather than the chaperone-like stabilizing role previously assigned to the DM proteins as a function in the class II-mediated Ag-presentation pathway (33, 34).

Coja-BG, -Lec, and -NK genes

The quail genes that are in synteny with the chicken B-G genes were tentatively designated as Coja-BG instead of Coja-G to avoid confusion with the quail class I gene, Coja-G. Of the eight Coja-BG loci, only three genes, BG4, BG6, and BG7, were predicted by Genscan analysis to have intact genomic structures consisting of 22, 15, and 16 exons, respectively, that would express...
FIGURE 4. Dot matrix analysis of the 180-kb Coja region vs itself (A) and between the Coja and chicken B-F/B-L regions (B). Diagonals indicate the regions where contiguous sequences conform to the stated parameters. Filled enclosures indicate the duplicated regions involving class I or class IIB genes.
transcripts of 5–6 kb (Fig. 5 and Table I). Gene orientation of BG6 and BG7 is the same as the chicken B-G gene whereas BG4 is in the opposite direction (Fig. 2B). Of the various coding exons, exons 1–3 and the last exon are well conserved with 87.6–89.8% nucleotide sequence identity. Exon 2 encodes the Ig variable-like region with 42% amino acid identity to both the human myelin oligodendrocyte glycoprotein and the human butyrophilin (Gen-Bank accession numbers Q61885 and U90548, respectively) (Fig. 3B). Exons 1 and 3 were predicted to encode the hydrophobic transmembrane domain by the SOSUI program. However, the internal exons between exon 3 and the last exon were not well conserved with ~65% nucleotide sequence difference among BG4, BG6, and BG7. All of these exons are composed of 21 nucleotides and they encode the α-helical coiled coil structure as predicted by the Chou-Fasman program (Fig. 5). The other five Coja-BG loci are apparently pseudogenes. Namely, BG1 has a nonsense mutation in exon 2, and BG8, BG5, BG3, and BG2 are cryptic genes lacking most exons (Table I and Fig. 5). Phylogenetic analysis of their Ig variable-like region domains (exon 2) suggests that the genetic distance from the chicken B-G gene is closer to Coja-BG6 than to BG4 and BG7 (Fig. 3B). Therefore, BG6 may be the quail orthologous gene for the chicken B-G, although the number of exons is different between these two genes, 10 for BG6 and 11 for B-G (35, 36).

Of the six Coja-Lec loci, four genes (Lec2, Lec3, Lec5, and Lec6) have four exons. Three of these genes (Lec3, Lec5, and Lec6) are pseudogenes because they lack an initiation codon in their coding regions. Lec1 and 4 are also pseudogenes with only single exons 3 and 4. In contrast, Lec2 has the entire 468-bp coding region and the highest nucleotide identity (93.0%) with the chicken B-Lec gene. The Lec2 gene encodes a C-type lectin-like protein deduced to consist of 156 aa with amino acid sequence similarities to the C-type lectin-like domains of a human lectin-like NK cell receptor protein (43% identity), human AICL (40% identity), mouse CLRF (42% identity), and a human early lymphocyte activation Ag CD69 (34% identity) (GenBank accession numbers AF133299, X96719, AF350410, and Z30426, respectively). All of the human genes encoding these proteins are located within the NK cell receptor region on the chromosome 12p13-p12 (37).

On the basis of their sequence identity and gene structure, the four Coja-NK loci (NK1 to NK4) all appear to encode proteins. The nucleotide sequence identities in the coding region of the Coja-NK genes range from 96.0 to 97.2%. The nucleotide sequence identities between each of the Coja-NK1–4 genes and the chicken B-NK gene range from 83.4 to 84.6%. The Coja-NK1–4 genes are composed of 136 aa, encoding a C-type lectin-like receptor with amino acid identity to the C-type lectin-like domain segments of a human NK cell receptor protein, NKR-P1A (33% identity) and a human cell adhesion molecule, CD209 (ICAM-3) (27% identity) (Gen-Bank accession numbers NP_002249 and NM_021155, respectively). The human genes encoding these proteins are located on the human chromosome 12p13-p12 within the NK cell receptor gene cluster and 19p13 within the *Mhc* paralogous genomic regions, respectively (37, 38).

The Coja-Lec2 and Coja-NK1–4 gene products have a low amino acid identity with each other (29%) but they both belong to the C-type lectin-like gene superfamily and they share seven conserved cysteine residues involved in disulfide bridges that may play a role in eliciting their receptor function. The seven cysteine residues are well conserved also in *LTL1, CLRF, AICL, CD69, NKR-P1A*, and *CD209* at the corresponding positions as seen in chicken B-NK (3). The Coja-Lec2 shares the C-type lectin-like domain with Coja-NK1–4, B-Lec, B-NK, and other immune-response-related genes in various species.

Other genes

The other quail genes, *TAPBPL, TAP1, TAP2*, and *RING3* (Figs. 2B and 4B), are found within orthologous positions of the chicken *Mhc* region (3) but in the class II gene region of the human and mouse *Mhc* (28, 39). Interestingly, the quail and chicken *TAP1* and *TAP2* genes reside within the class I region, as similarly observed in the zebrafish, pufferfish, and Xenopus *Mhc* regions (40–42).

The quail *TAPBPL* or *Tapasin* gene, known as the TAP-binding protein gene (43), was found within a 6-kb segment between the DCB1 and DBB1 genes (Fig. 2B). It is homologous to the chicken, human, and mouse *Tapasin* genes with 90, 34, and 33% amino acid identities, respectively. The quail *TAP1* and *TAP2* genes that belong to the ATP-binding cassette *Mhc* class 1 Ag-transporter gene family (44) were identified within a 10-kb segment between the *D2* and *B1* genes (Fig. 2B). The amino acid sequence identities of the quail *TAP1* and *TAP2* proteins were 91, 45, and 43% with the chicken, human, and mouse *TAP1* and *TAP2*, and 41% with their *TAP2* proteins, respectively (23). The amino acid sequence of *TAP2* in contig 1 has two amino acid substitutions (Thr/Lys and Asp/Asn).
Table IV. Diversity of 18 peptide binding sites in Coja and Gado class IIB genes

<table>
<thead>
<tr>
<th>Locus</th>
<th>-DAB1</th>
<th>-DAB2</th>
<th>-DBB1</th>
<th>-DBB2</th>
<th>-DBB3</th>
<th>-DBB4</th>
<th>-DBB5</th>
<th>BLB1</th>
<th>BLB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAB1</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDB1</td>
<td>14</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEB1</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td></td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFB1</td>
<td>12</td>
<td>8</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGB1</td>
<td>14</td>
<td>14</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-LB1</td>
<td>14</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>16</td>
<td>12</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-LB2</td>
<td>12</td>
<td>12</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>10</td>
<td>16</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

a Arabic numerals indicate the numbers of divergent amino acid residues on the 18 peptide-binding sites between the class IIB loci. Bold numbers indicate >50% divergent amino acids.

Ala/Gln at amino acid positions 328 and 329, respectively) (Fig. 1) when compared with TAP2 in contig 2/3 (Ref. 15, GenBank accession number AB007795, and Fig. 1) due to two single nucleotide polymorphisms between the two different haplotypes. The genomic and deduced amino acid sequences of TAPBP2, TAP1, and TAP2 in contig 1–3 had no structural defects that could be deleterious to Mhc class I or class II Ag presentation.

The RING3 gene that encodes a nuclear serine-threonine kinase protein and has the bromodomain (45) was identified within a 14-kb segment between the DAB1 and DMA1 genes (Fig. 2B). It is homologous to the chicken, human, and mouse RING3 genes with 99, 69, and 68% amino acid identity, respectively. Exon positions were determined by sequence alignment with the quail and chicken RING3 genomic sequences. Consequently, quail RING3 was found to consist of 11 exons in contrast to the 12 exons in the human and mouse RING3 genes. This difference can be explained by the absence of a quail exon that would correspond to exon 1 in the human and the mouse and transcribe the 5′ untranslated region. The quail RING3 protein (735 aa) and chicken RING3 protein (733 aa) are smaller than the human (798 aa) and mouse (801 aa) RING3 proteins. The biological function of RING3 remains to be established although its gene is always linked to the Mhc of all the jawed vertebrates so far examined (29).

In human and mouse, the C4 gene that encodes the complement factor C4 resides in the class III region that is located between the class II and class I regions (28, 39). In quail, the C4 gene was found within a 17-kb segment between the E and histone-H3 genes (Fig. 2B). The quail C4 has 38 exons with 91.7% nucleotide and 91.8% amino acid sequence identity with the chicken C4. The translated quail C4 gene had ~35% amino acid sequence identity with the human and mouse C4A and C4B proteins. The highest similarity was observed within the domain regions of the α2-macroglobulin N-terminal family, the anaphylotoxin-like domain, the α2-macroglobulin family, and the netrin C-terminal (NTR/C345C) domain, which are essential for interacting with the C3, C4, and C5 proteins (data not shown).

The quail and chicken histone-H3-like genes, located at the terminal end of the class I genes regions, are pseudogenes that lack a large part of their coding regions. As with the Gado B region (3), there were no class IIa and LMP2/7 genes found within the Coja Mhc region.

Dot-matrix analysis

The patterns of regional genomic duplication were examined by dot-matrix analysis of the 180-kb Coja sequence (Fig. 4A). The analysis revealed that the class IIB region has >22 pairs of duplicated segments ranging in size from 1 to 20 kb. This striking observation prompted us to look more closely into each of these segments. Five class IIB genes (DGB1, DFB1, DEB1, DDB1, and DCB1) are oriented in the same direction and two upstream 20-kb segments have a high degree of similarity. The 20-kb upstream segment with DCB1 (BG2, Nk1, BG1, Lec2, and DCB1) has

FIGURE 5. Exon-intron and structural organization of the quail and chicken BG genes. TMR indicates the transmembrane regions. Striped and filled boxes encode the Ig variable regions and α-helical coiled coils structure regions, respectively. Coja-BG2, -BG3, -BG5, and -BG8 are pseudogenes containing only exons 1, 2, and 3.
95.7% nucleotide similarity with the 20-kb upstream segment carrying DGB1 (BG5, NK4, BG4, Lec6, and DGB1). Therefore, there are five class IIB-linked homologous segments that share a unique combination of three genes that are members of the NK, BG, and Lec multigene families, such as the DCB1 segment (BG2, NK1, BG1, Lec2, and DCB1), the DDB1 segment (NK2, Lec3, and DDB1), the DEB1 segment (Lec4 and DEB1), the DBF1 segment (BG3, NK3, Lec5, and DBF1) and the DGB1 segment (BG5, NK4, BG4, Lec6, and DGB1). Furthermore, the DBF1/DEB1 unit (DBF1, LINE, Lec4, and DEB1) resembles the DDB1/DEB1 unit (DBB1, Lec1, LINE, and DAB1), although they are in the opposite direction to each other (Fig. 2B). Thus, successive cis- or trans-segmental duplications of this basic unit appear to have produced the present Coja class IIB region. No similar segmental duplications were recognized in the quail class I or the chicken class I or class II region by dot matrix analyses. However, dot matrix of the Coja and the B-F/B-L regions revealed that the Coja class IIB duplicated units converged at the single chicken Mhc class IIB unit (B-G, B-NK, B-Lec and B-LB1) (Fig. 4B). The RING3 to Coja-E region is well conserved between two species with 84% nucleotide sequence identity, when the differences for the genes (53,287 bp), exons (18,090 bp) and introns (35,197 bp) were 13.9, 10.8, and 16.2%, respectively (Fig. 6). The average amino acid sequence divergence for the coding regions (6030 aa) was 14.7%. The range of diversities between the 13 genes for each of the analyses can be seen in Fig. 6. The amino acid divergences varied considerably between the quail and the chicken depending on the gene products (Fig. 6D). In contrast, the nucleotide diversities were relatively uniform across the entire gene sequence for all 13 genes (Fig. 6, A–C). In this regard, the amino acid sequences of the quail BG6, NK1, DCB1, and B1 proteins, which fulfill immunological functions as cell surface molecules, were more divergent than in the other proteins. For example, the amino acid divergence of NK1 (31.3%) was twofold higher than the average divergence (14.7%), and BG6, DCB1, and B1 were also highly divergent (22.0, 20.1, and 23.7%, respectively). DMA1, DMB1, and DMB2 displayed moderate divergence that was close to the average divergence between quail and chicken (17.5, 14.3, and 16.2%, respectively). Lec2 was less divergent than other cell surface molecules (10.3%). TAPBPL, TAP1, and TAP2 were more conserved with similar diversities of 10.2, 9.3, and 9.7%, respectively, than the other proteins. RING3 was the most conserved protein sequence with the lowest amino acid diversity of 1.2% between quail and chicken, supporting the previous finding that the RING3 is evolutionarily stable and highly conserved from yeast to human.

It is noteworthy that BG6 and NK1 revealed a surprisingly high degree of amino acid diversity. The diversity between these loci in quail and chicken is equivalent to or higher than that at the class I (B1) and class II (DCB1) loci. This is surprising because class I and class II are among the most polymorphic loci in vertebrates. This observation suggests that BG6 and NK1 may also be under

**FIGURE 6.** Nucleotide and amino acid diversities between quail and chicken for the separate categories of gene regions (A), intronic regions (B), exonic regions (C), and amino acid sequences (D) in 13 representative genes. Coja-BG6 was chosen in this analysis because it was the most homologous to the chicken B-G gene. Coja-Lec2, Coja-NK1, and Coja-DCB1 were chosen as representatives of their respective Coja multigene families. Comparison of Coja-BG6 and chicken B-G was performed in the exon 1–3 regions. Other comparisons included the entire gene regions. Gray lines indicate average percentages.
selection for diversity. Therefore, not only the class I B1 and class II DCB1, but also the BG6 and NK1 appear to undergo selective pressure to maintain genetic polymorphisms. The other expressed NK and BG genes also encode a high degree of amino acid divergence (30.9% for NK2, 33.1% for NK3, 30.9% for NK4, 24.2% for BG4, and 24.2% for BG7) when compared with the corresponding chicken B-NK and B-G proteins.

The percentage nucleotide divergence of the intronic regions between quail and chicken was relatively uniform at 16.2% for the 13 genes on average (Fig. 6B). The separation time (T) between the quail (Coja) and chicken (B-F/B-L) Mhc was calculated as 81 ± 1.9 million years ago by using the equation $d = 2\mu T$, where $d$ is the average nucleotide distance (Kimura 2 parameter; 0.162) between the intronic regions of the two species and $\mu$ is the mutation rate determined for introns and pseudogenes in primates ($1.0 \times 10^{-9}$) (46).

**Discussion**

The genomic sequence, gene structure, and organization of the 180-kb quail Mhc Coja region was determined for one contig and compared with that of the chicken B-F/B-L region. The contig 1 represented a single Coja haplotype that contained 7 class I genes and 10 class II genes (Fig. 1). Contig 2–5 represented the other haplotype with 12 class I genes (23) and 9 class IIB genes, excluding the DMB and DMA loci. Collectively, at least 14 class I and >10 isotypically different class IIB loci that are distinguished by nucleotide sequence identity and gene location can be present in the quail diploid genomes of a single individual. The numbers of quail Mhc class I and class IIB genes are consistent with those predicted by Southern hybridization (23, 24). Therefore, we believe that we have identified most if not all of the class I and class II genes in the quail. Other birds such as the great reed warbler and the willow warbler were found to contain 13–17 and 25–30 class I bands per individual by Southern hybridization, respectively (47). Furthermore, the chicken minimal essential Mhc characterized by its smaller size and simplicity (3) cannot be generalized to other avian species.

The structural differences between the Mhc regions of the quail and chicken may be explained by segmental duplication of two distinct duplication units (Fig. 4). For instance, the genomic segment consisting of the Coja-class IIB, Lec, BG, NK, and BG gene family members has been duplicated at least five times along with some deletion events to form part of the present-day Coja class IIB region. Similarly, repeated gene duplications of the essential class I segment have contributed to the main organization of the Coja class I region (Fig. 4). Phylogenetic analysis of the quail and chicken Mhc class I and class II genes showed that these gene sequences were more closely related within species than between species, suggesting that the quail class I and class II genes were generated after the separation of the quail and chicken from their common ancestor. The genomic structure of the Coja Mhc region was probably shaped by duplication events in response to environmental conditions and pathogens that the quail encountered after its divergence from the chicken. For example, the chicken is a nonmigratory bird originating from Southeastern Asia while the quail is a migratory bird originating from Northern and Southern Asia and flying only short distances at a time. Therefore, the quail immune response system may have been stimulated by a larger variety of environmental pathogens during its migratory history, promoting an increase in the number of class I and class IIB loci and/or the degree of genetic polymorphism. A similar explanation may be applied to other migratory birds such as the warbler (47). In addition, the chicken may have lost or translocated several Mhc loci and gene isotypes after its separation from the quail lineage and during its domestication and/or expansion of Rp-Y, resulting in the emergence of the minimal essential Mhc (32, 48).

To trace the evolutionary process of the quail Mhc class II region, phylogenetic analyses were performed using the BG, NK, and Lec amino acid sequences that were derived from the nucleotide-coding sequences (Fig. 3). However, these analyses could not identify the progenitor gene nor elucidate the detailed duplication processes, probably because large-scale insertions/deletions have taken place frequently in the Coja class II regions after the regional duplication events (Fig. 4). Generally, our findings support the “birth and death” model for the evolution of vertebrate multigene families (49, 50). In addition, eight LINE and seventeen LTR sequences were identified in the 180-kb Coja region, but no LINE and LTR sequences were present in the 92-kb Gado Mhc region (Fig. 2D). In the human Mhc region, the density of LTR is significantly higher around the duplicated regions. For example, the Mhc class I polypeptide-related sequence B (MICB) to HLA-C region has an LTR content of 23% on average, whereas nonduplicated regions have an LTR content of 8.2% on average. Indeed, both ends of the HLA class I-duplicated units contain LTR sequences (51, 52), implying that there were retrotransposon-mediated segmental genome duplications within the Mhc regions (53, 54). Therefore, the presence and absence of the retrotransposons in the quail Mhc class II region and the chicken Mhc class I and class II regions, respectively, is consistent with a model of retrotransposon-driven segmental genome duplication.

Of the class I and class IIB expressed loci that were identified in contig 1 of the quail Mhc region, D1, D2, DCB1, DDB1, DEB1, and DGB1 appear to be nonclassical class I or class II loci in terms of a weak expression and limited tissue-specificity as previously observed for the B-FA2 and B-LB1 genes in the chicken Mhc region. Although the function of the DMA1, DMB1, and DMZ2 genes are not known, they are unlikely to be the major class II loci. In contrast, the expression of B1, E, DAB1, and DDB1 is very high in a large variety of tissues, as similarly observed for the B-FA1 and B-LB2 genes in the chicken. Therefore, these genes are probably the major or classical class I and class II loci (3, 32). It is likely that the classical class I and class II genes acquired their locus-specific function gradually by repeated nonsynonymous substitutions after their generation by duplication. The amino acid diversities between the quail and chicken Mhc class I and class IIB loci were found to be as high as 20% and the maintenance or progression of genetic polymorphisms within these two species is probably favored by a balancing or overdominant selection. The TAPBPL, TAP1, TAP2, DMA1, DMB1, and DMB2 genes were also found to have relatively high degrees of amino acid diversity, ranging from 9.3 to 17.4% (Fig. 6D). The amino acid divergences observed in the Ag-processing pathway-associated transporters or chaperones are much higher than that in RING3 (1.2%), which seemingly is a genetically conserved housekeeping or regulatory gene (45). Therefore, it may be speculated that the quail Mhc class I and class II genes and its Mhc Ag-processing genes have coevolved through various genetic interactions in pursuit of efficient allelic combinations that are necessary for the establishment and maintenance of a sophisticated immune response system.

The peptide binding sites deduced from the amino acid sequences of the quail Mhc class I and class IIB genes are highly variable (63% amino acid diversity) among the class I and class IIB proteins (Tables III and IV, and Ref. 23). In contrast, the amino acid diversities between the chicken’s class I proteins (B-FA1 and B-FA2) and between the class II proteins (B-LB1 and B-LB2) were much less than between the corresponding quail class I and class II proteins (Table IV). This suggests that the Ag-presentation repertoire in the quail is larger and better adapted to environmental
intronic regions (14) which is similar to the average divergences in other chromosomal B-loci in xenotransplantation.

Based on the average nucleotide diversity that have been maintained under positive selection (Fig. 6D). The average nucleotide sequence divergence of intronic regions in the quail as well as in other birds. Such comparable genotypic studies among avian species will give us more information about the genetic, immunological, and evolutionary differences that exist between avian genomes.

References


2. Guillermot, F., A. Billault, O. Pourquier, G. Behar, A. Chausse, R. Zoorob, G. Kreibich, and C. Auffray. 1988. A molecular map of the chicken major histocompatibility complex: the class II genes are closely linked to the class I genes and the nuclear organizer. EMBO J. 7:2775.


major histocompatibility complex (MhcCoja) class I gene region. Immunogenetics 49:384.


