6-2003

Comparative Genome Organization of Human, Murine, and Feline MHC Class II Region

Naoya Yuhki
National Cancer Institute at Frederick

Thomas W. Beck
National Cancer Institute at Frederick

Robert M. Stephens
National Cancer Institute at Frederick

Yoko Nishigaki
National Cancer Institute at Frederick

Kymberly Newmann

See next page for additional authors

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

Part of the Biodiversity Commons, Genetics and Genomics Commons, and the Zoology Commons

NSUWorks Citation

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Authors
Naoya Yuhki, Thomas W. Beck, Robert M. Stephens, Yoko Nishigaki, Kymberly Newmann, and Stephen J. O'Brien

This article is available at NSUWorks: https://nsuworks.nova.edu/cnso_bio_facarticles/791
Comparative Genome Organization of Human, Murine, and Feline MHC Class II Region

Naoya Yuhki,1,4 Thomas Beck,2 Robert M. Stephens,3 Yoko Nishigaki,2 Kymberly Newmann,1 and Stephen J. O’Brien1,4

1Laboratory of Genomic Diversity, National Cancer Institute–Frederick, Frederick, Maryland 21702, USA; 2Intramural Research Support Program (IRSP), SAIC-Frederick, Frederick, Maryland 21702, USA; 3Advanced Biomedical Computing Center, IRSP, SAIC-Frederick, Frederick, Maryland 21702, USA

To study comparative molecular dynamics in the genesis of the major histocompatibility complex (MHC), we determined a complete nucleotide sequence spanning 758,291 bp of the domestic cat (*Felis catus*) extended and classical class II region. The feline class II MHC includes 44 genes (31 predicted to be expressed) which display DNA sequence homology and ordered gene syntenies with human HLA and mouse H2 in extended class II and centromere proximal regions (DM to DO) of the classical class II region. However, remarkable genomic alterations including gene gain and loss plus size differentials of 250 kb are evident in comparisons of the cat class II with those of human and mouse. The cat MHC lacks the entire BQ region and retains only relict pseudogene homologs of DR genes, compensated by expansion and reorganization of seven modern DR genes. Repetitive gene families within the feline MHC comprise 35% of the feline MHC with very different density and abundance of GC levels, SINES, LINES, STRs, and retro-elements from the same repeats in human and mouse MHC. Comparison of the feline MHC with the murine and human MHC offers a detailed view of the consequences of genome organization in three mammalian lineages.

The vertebrate major histocompatibility complex (MHC) offers an unusual opportunity for comparative genomics, as it consists of a chromosomally linked community of over 100 expressed genes, nearly half having a defined role in immune defenses. A complete sequence of human HLA (3.6 Mb), chicken B-locus (92 kb), and a partial mouse H2 sequence (4.0 Mb) have been determined (Kaufman et al. 1999; The MHC Sequencing Consortium 1999; Shihina et al. 1999; Stephens et al. 1999; Yu et al. 2000; Takada et al. 2003; L. Rowen, pers. comm.). Initial comparisons reveal remarkable differences in gene inclusion, gene order, and sequence divergence, likely driven by natural selective pressures of infectious disease outbreaks in the ancestors of these species (Doherty and Zinkernagel 1976; Klein 1986; Farham and Ohta 1996; Zinkernagel 1996; Hughes and Yeager 1998; Carrington et al. 1999). A deep understanding of the immunological functions, disease influences, evolutionary constraints, and driving forces of MHC adaptation in several mammalian species provides a fertile venue to discern and interpret adaptive events which predate living vertebrate genomes.

HLA, the human MHC, is composed of 224 tightly linked genes, including 128 expressed genes, 96 pseudogenes, and repetitive elements (SINES, LINES, LTRs, and STRs) that comprise nearly 50% of the sequence (The MHC Sequencing Consortium 1999; Shihina et al. 1999). The mouse MHC, H2, includes 120 expressed genes encoding functions similar to those ascribed to human MHC genes plus various pseudogenes and repeat sequences (Klein 1986; Stephens et al. 1999; Yu et al. 2000; Takada et al. 2003; L. Rowen, pers. comm.).

*Corresponding authors.
E-MAIL yuhki@ncifcrf.gov; FAX (301) 846-6327.
E-MAIL obrien@ncifcrf.gov; FAX (301) 846-1909.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.976103. Article published online before print in May 2003.

Both human and mouse MHC show linked segments of five subregions named for the individual gene families and ordered as follows from the centromere (size of the human sub-region): (1) extended class II (300 kb), (2) classical class II (850 kb), (3) class III (700 kb), (4) classical class I (1800 kb), and (5) extended class I (1000 kb). Within each MHC genomic region, however, there are significant differences between the two species. For example, class I homologs within human or within mice are more orthologous to each other than they are to homologs between the two species. In addition, the many pseudogenes are found predominantly in class I and II regions and are virtually absent in the MHC central (class III) region.

The chicken MHC, B-locus, presents a “minimal essential MHC” disposition extending 92 kb and including just 19 functional genes, 12 of which share homologs in the human and mouse MHC (class Iβ, tapasin, class Iβ, Ring 3, DMA, DMB, DMβ, class I, Tα1, Tα2, class I, C4; Kaufman 1999). The order of the chicken B-locus gene segments is different from those of human and mouse, and the relative compactness of the chicken MHC (~5% of the human MHC) raises the possibility that the ancestral vertebrate MHC was smaller (like the chicken’s) and was expanded during the mammalian radiations.

The MHC class II region is of particular interest because very many of the ~40 genes included are specifically involved in antigen processing and presentation to the T-cell receptors. The class II region shows evidence of rapid adaptive duplications and deletions as well as allele and haplotype polymorphism within and between human and mouse species. Here we present the complete 758,291-bp sequence of the feline MHC class II region, including the extended and classical class II regions. The cat was chosen to provide comparative MHC sequence from a third order of placental mammals, Carnivora, to augment comparative genomic analysis in domes-
tic cats (O’Brien et al. 1997, 1999; Murphy et al. 2000) and the Felidae Family (Johnson and O’Brien 1997; Pecon-Slattery and O’Brien 1998) as well as to contribute a powerful genomic tool to develop immunogenetic insight into feline hereditary (http://www.angis.org.au/Database/BIRX) and infectious disease models, for example, feline leukemia virus (Hoover and Mullins 1991), feline immunodeficiency virus (Pederson et al. 1987), and feline infections peritonitis virus (Andrew 2000).

RESULTS

A composite 758,291-bp sequence was obtained from five PAC (f2, f3, f11, f27, and g7) and four BAC clones (186b21, 102h1, 463h11, and 160a17) which spanned the extended and classical class II region of the domestic cat. Forty-four coding genes were identified by GENESCAN and BLASTP algorithms from the genes HSET (centromeric extreme of extended class II) to BTLII (telomeric of class II region). A description of the function and characteristics of each class II region gene is given in Table 1. A molecular sequence identity map which compares the feline class II sequence with the human sequence to identify gene homologs, exons, introns, and other genomic features is presented in Figure 1. A comparative alignment of genes found in human, cat, and mouse class II regions is illustrated in Figure 2.

Comparison of the gene organization of the MHC class II region between the mouse, domestic cat, and human indicates size differences of about 250 kb among these three species (Fig. 2). The sizes of the MHC class II region are 495 kb, 758 kb, and 998 kb in the mouse, domestic cat, and human, respectively. These size differences appear to reflect both the presence/absence of genes plus moderate differences in gene density in each MHC. The observed gene densities were one gene every 13.7 kb, 17.2 kb, and 18.1 kb in the mouse, domestic cat, and human MHC class II regions, respectively.

The feline extended class II region includes 16 genes (Table 1, Fig. 2). The order of the 16 feline genes is identical to that in human, indicating retention of an ancestral gene order for these genes. The cat MHC contains a single gene RPS28 (40s ribosomal protein S28) between the HSET and DAXX loci, which is absent in human. Three pseudogenes (RPL12-L, RPL35A-L, and BING3) and one functional gene (APT-2), all absent in cat, occupy this same region in the human MHC. Both segments are absent in mouse, perhaps indicating a Primate- and Felidae-specific gene segment transposition into this region. The mouse extended class II region is also missing a segment that specifies [TATSF1-ZNFL] in human and TATSF1 alone in cat. In that position, the murine MHC includes a stretch of two class I region genes [H2K2, H2K1]. The feline classical class II region contains 28 genes. Of these, 19 are homologous to human MHC homologs whereas three genes, namely GAPDH (glyceraldehyde 3′-phosphate dehydrogenase), SURF3 (L7a ribosomal protein like gene), and v-ATPase (vacular ATP synthetase) are transposed pseudogenes related to genes located on human chromosomes 12p13 (GAPDH), 9q34 (SURF3), and 3p, 5, 6p21.3 MHC class III region, and 8q22.3 (v-ATPase). In addition, the cat class II region includes three LINE1 reverse transcriptase genes (Smit 1999) and three open reading frames (X) of unknown function (Table 1). Four multigene segments of the human MHC are missing in the feline MHC (labeled A–D in Fig. 2): (A) [RPL12-L, RPL35A-L, APT2, BING3], (B) [COL11A2, DPA2, DPB1, RPL32-L, DPA1], (C) [HLA-Z1, RING14, IFP2, RING8], and (D) the entire DQ gene family region [DQB2, DQA2, DQB3, COX3CL, GLN-8R1, DQB1, DQA1]. The feline DR region consists of four DRB genes and three DRA genes compared to the human’s one DRA and two to five DRB genes. The feline DR region genes do not display definitive sequence orthology to the human counterparts, and their gene order is rearranged to suggest a history of duplication and inversion events during Carnivora evolutionary history (see below).

The mouse extended class II region displays a high degree of gene sequence and syntenic homology to the human and cat regions (Fig. 2). The single exception is the placement of H-2K1 and H-2K2 class I homologs in a position occupied by two pseudogenes [TATSF1 and ZNFL-1] in human and one pseudogene, TAT- SF1, in the cat (see bracket E in Fig. 2). The mouse classical class II region shows clear gene sequence and order homology to the human/cat region spanning 11 genes (from Pb to Tap2 except Mb2). Beyond Tap2 however, there is very limited homology except for distant similarity of Eb1, Eb2, Ea to DR and Ab1, Ab1, Ab2 to D0/DQ region genes of human and cat. The divergence of mouse class II sequence is striking in the face of more parallel cat/human comparisons and may reflect macro-evolutionarily reorganization/divergence events in the history of rodent species.

Dot plot analyses (Sonnhhammer and Durbin 1995) between the domestic cat sequence and a corresponding sequence of human class II region (997,836 bp) showed clear-cut conserved and nonconserved regions between the two species’ MHC class II sequences (Fig. 3A). Two long stretches of conserved sequences between cat and human class II region were found; one in the extended class II region (~250 kb) and a second in the classical class II region (~220 kb). Upon close inspection using PILFsnake (Fig. 1; Schwarz et al. 2000), the two highly conserved stretches were shown to arise from clustered exonic sequences. In the first, retinoic acid X receptor α gene (RXRB) and collagen type 11 a2 domain (COL11A2) coding gene have 70 predicted exons in a 40-kb nucleotide sequence in the extended class II region with more than 75% sequence identity between cat and human sequences (Fig. 1). Similarly, 31 conserved exons included in four genes (LMP2, TAP1, LMP7, TAP2) were clustered within a 30-kb segment of the classical class II region (Fig. 1). These two conserved gene stretches were interrupted by a nonconserved 100 kb of DP subregion. The 100-kb stretch of the human sequence which encodes DQ subregion was completely missing in the cat sequence.

A dot plot analysis (Sonnhammer and Durbin 1995) of human (234 kb) and cat DR subregions (266 kb) detected at least three cat DRB homology segments spanning 15- to 30-kb sequences and three cat DRA signals spanning 15-kb sequences (Fig. 3B). Self-dot plot analysis of cat DR subregion revealed three 15-kb DRB gene repeats, three 15-kb DRA gene repeats, plus two 30-kb repeats including both DRB and DRA genes (Fig. 3C). These homology signals plus the observation that human DRB paralogs and feline DRB paralogs have equivalent amino acid sequence similarity within the two species (88%–94% in cat DRB1,3,4 vs. 88 to 93% in human DRB1,3,4,5), as is apparent between the species (80%–82% cat vs. human), would suggest the occurrence of duplication subsequent to split of primate and carnivore and reordering of the DRB and DRA gene in the two ordinal lineages. A parsimonious interpretation for the development of the feline DR gene order is illustrated in Figure 4.

Comparative analysis of interspersed repeat sequences in the MHC class II regions among human, domestic cat, and mouse indicated that five types of interspersed repeats,
namely SINES, LINES, LTR transposons, short tandem repeats (STRs), and DNA transposons were common to all three species. There were however notable differences in the proportion of the repeat sequences in the three species’ MHC class II regions (Table 2). Human ALU-SINE repeats (Rowold and Herrera 2000) occupy threefold greater sequence density compared to human LINE repeats in the extended class II density region (27% for SINES and 6.8% for LINES). In contrast, hu-

### Table 1. Genes and Predicted Protein Products From HSET to BTLII of the Cat MHC Class II Region

<table>
<thead>
<tr>
<th>Gene</th>
<th>Functional/physiological properties/other name/structure</th>
<th>Size (bp)</th>
<th>Predicted protein size (aa)</th>
<th>No. of exons</th>
<th>Distance to next gene (bp)</th>
<th>% aa sequence similarity Human Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extended Class II Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSET</td>
<td>Kinesin motor protein-related, kifc, KNSL2, kinesin-like protein 2</td>
<td>&gt;13226</td>
<td>&gt;657</td>
<td>&gt;10</td>
<td>12643</td>
<td>91 83</td>
</tr>
<tr>
<td>RPS28</td>
<td>Component of 528 of ribosome</td>
<td>1861</td>
<td>103</td>
<td>2</td>
<td>46489</td>
<td>55 55</td>
</tr>
<tr>
<td>DAXX</td>
<td>Fas binding protein, apoptosis through c-Jun</td>
<td>3197</td>
<td>748</td>
<td>7</td>
<td>1754</td>
<td>91 80</td>
</tr>
<tr>
<td>BING1</td>
<td>Zinc finger domain</td>
<td>1907</td>
<td>635</td>
<td>1</td>
<td>959</td>
<td>95 88</td>
</tr>
<tr>
<td>TAPASIN</td>
<td>Ig superfamily, associate with class I antigen during peptide loading</td>
<td>11062</td>
<td>488</td>
<td>9</td>
<td>2207</td>
<td>85 75</td>
</tr>
<tr>
<td>RGL2</td>
<td>Rat guanine nucleotide dissociation factor</td>
<td>6658</td>
<td>664</td>
<td>15</td>
<td>3121</td>
<td>94 92</td>
</tr>
<tr>
<td>BING4</td>
<td>WD40 domain</td>
<td>7638</td>
<td>612</td>
<td>15</td>
<td>709</td>
<td>93 87</td>
</tr>
<tr>
<td>B3GALT4</td>
<td>β 1,3-galactosidase-4</td>
<td>1152</td>
<td>378</td>
<td>2</td>
<td>1000</td>
<td>86 81</td>
</tr>
<tr>
<td>RPS18</td>
<td>Component of S18 of ribosome</td>
<td>5011</td>
<td>152</td>
<td>6</td>
<td>438</td>
<td>100 100</td>
</tr>
<tr>
<td>ARE1</td>
<td>Yeast sac2 homolog, suppressor of actin mutant 2, Sacm2, coiled coil structure</td>
<td>15145</td>
<td>702</td>
<td>17</td>
<td>28038</td>
<td>72 99</td>
</tr>
<tr>
<td>TAT-SF1</td>
<td>HIV1 transcriptional elongation factor, pseudogene</td>
<td>ND</td>
<td>ND</td>
<td>1</td>
<td>ND</td>
<td>ND ND</td>
</tr>
<tr>
<td>RING1</td>
<td>Ring finger protein, polycym group protein</td>
<td>2351</td>
<td>307</td>
<td>6</td>
<td>4129</td>
<td>99 98</td>
</tr>
<tr>
<td>KE6</td>
<td>Steroid dehydrogenase-like protein (estradiol 17β-dehydrogenase B)</td>
<td>ND</td>
<td>149</td>
<td>4</td>
<td>ND</td>
<td>ND ND</td>
</tr>
<tr>
<td>KE4</td>
<td>Transmembrane protein with histidine-rich charge clusters</td>
<td>4516</td>
<td>193</td>
<td>4</td>
<td>858</td>
<td>83 82</td>
</tr>
<tr>
<td>RXRβ</td>
<td>Retinoid X receptor, β</td>
<td>5559</td>
<td>475</td>
<td>10</td>
<td>2515</td>
<td>100 97</td>
</tr>
<tr>
<td><strong>Classical Class II Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRβ</td>
<td>Class II antigen, β chain, pseudogene</td>
<td>22891</td>
<td>317</td>
<td>5</td>
<td>2597</td>
<td>60 ND</td>
</tr>
<tr>
<td>DPA</td>
<td>Class II antigen, α chain, pseudogene</td>
<td>8678</td>
<td>235</td>
<td>5</td>
<td>5075</td>
<td>74 ND</td>
</tr>
<tr>
<td>X</td>
<td>Unknown function</td>
<td>9161</td>
<td>205</td>
<td>2</td>
<td>10553</td>
<td>ND ND</td>
</tr>
<tr>
<td>DNA</td>
<td>DOA, heterodime with DOB in pre-B cells, peptide loading for class II antigen at low PH</td>
<td>9622</td>
<td>265</td>
<td>5</td>
<td>15757</td>
<td>86 77</td>
</tr>
<tr>
<td>RING3</td>
<td>Nuclear-localized, serine-threonine kinase, frg-1</td>
<td>13809</td>
<td>1093</td>
<td>17</td>
<td>10773</td>
<td>98 95</td>
</tr>
<tr>
<td>DMA</td>
<td>Nonclassical class II antigen, α chain, peptide loading for class II antigen</td>
<td>3926</td>
<td>260</td>
<td>5</td>
<td>9876</td>
<td>86 81</td>
</tr>
<tr>
<td>DMB</td>
<td>Nonclassical class II antigen, β chain, heterodimer with DNA, peptide loading for class II antigen</td>
<td>9397</td>
<td>279</td>
<td>5</td>
<td>16548</td>
<td>82 70 (Mb1), 71 (Mb2)</td>
</tr>
<tr>
<td>V-ATPase</td>
<td>Vacuolar ATPase, pseudogene</td>
<td>9952</td>
<td>229</td>
<td>4</td>
<td>21688</td>
<td>70 70</td>
</tr>
<tr>
<td>LINE1</td>
<td>Reverse transcriptase homolog</td>
<td>464</td>
<td>154</td>
<td>1</td>
<td>15606</td>
<td>59 48</td>
</tr>
<tr>
<td>LMP2</td>
<td>Proteosome subunit to cleave peptides for class I antigen</td>
<td>3057</td>
<td>196</td>
<td>5</td>
<td>2528</td>
<td>92 91</td>
</tr>
<tr>
<td>TAP1</td>
<td>Transporter for antigen processing</td>
<td>7279</td>
<td>713</td>
<td>11</td>
<td>1822</td>
<td>82 75</td>
</tr>
<tr>
<td>LMP7</td>
<td>Transporter subunit to cleave peptides for class I antigen</td>
<td>2875</td>
<td>276</td>
<td>5</td>
<td>2591</td>
<td>82 85</td>
</tr>
<tr>
<td>TAP2</td>
<td>Transporter for antigen processing</td>
<td>9852</td>
<td>702</td>
<td>11</td>
<td>3451</td>
<td>79 77</td>
</tr>
<tr>
<td>LINE1</td>
<td>Reverse transcriptase homolog</td>
<td>1039</td>
<td>198</td>
<td>2</td>
<td>5980</td>
<td>50 45</td>
</tr>
<tr>
<td>DOI</td>
<td>Nonclassical class II antigen, β chain, H2-IAB2 in mouse</td>
<td>18247</td>
<td>209</td>
<td>4</td>
<td>1119</td>
<td>72 67</td>
</tr>
<tr>
<td>LINE1</td>
<td>Reverse transcriptase homolog</td>
<td>18237</td>
<td>222</td>
<td>4</td>
<td>14603</td>
<td>70 68</td>
</tr>
<tr>
<td>DRβ4</td>
<td>Class II antigen, β chain</td>
<td>12470</td>
<td>266</td>
<td>5</td>
<td>33292</td>
<td>81 (DRβ3) 78 (Eb1)</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde phosphate dehydrogenase, pseudogene</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND ND</td>
</tr>
<tr>
<td>DRβ1</td>
<td>Class II antigen, β chain</td>
<td>10009</td>
<td>266</td>
<td>5</td>
<td>14668</td>
<td>81 (DRβ3) 78 (Eb1)</td>
</tr>
<tr>
<td>DRA1</td>
<td>Class II antigen, α chain</td>
<td>2684</td>
<td>254</td>
<td>5</td>
<td>40740</td>
<td>91 81</td>
</tr>
<tr>
<td>SURF3</td>
<td>L7a ribosomal protein, pseudogene</td>
<td>ND</td>
<td>ND</td>
<td>1</td>
<td>inside intron 1 of DRβ2</td>
<td>ND ND</td>
</tr>
<tr>
<td>DRβ2</td>
<td>Class II antigen, β chain, intron 1 &amp; exon 2 gene fragment, pseudogene</td>
<td>ND</td>
<td>89</td>
<td>1</td>
<td>ND</td>
<td>ND ND</td>
</tr>
<tr>
<td>DRβ3</td>
<td>Class II antigen, β chain</td>
<td>9680</td>
<td>294</td>
<td>5</td>
<td>10790</td>
<td>81 (DRβ3) 80 (Eb1)</td>
</tr>
<tr>
<td>X</td>
<td>Unknown function</td>
<td>179</td>
<td>266</td>
<td>1</td>
<td>2666</td>
<td>ND ND</td>
</tr>
<tr>
<td>DRA2</td>
<td>Class II antigen, α chain</td>
<td>2684</td>
<td>254</td>
<td>5</td>
<td>24414</td>
<td>91 81</td>
</tr>
<tr>
<td>X</td>
<td>Unknown function</td>
<td>176</td>
<td>58</td>
<td>1</td>
<td>6120</td>
<td>ND ND</td>
</tr>
<tr>
<td>DRA3</td>
<td>Class II antigen, α chain</td>
<td>3132</td>
<td>254</td>
<td>5</td>
<td>19743</td>
<td>92 83</td>
</tr>
<tr>
<td>BTLII</td>
<td>Butyrophilin-like gene, immunoglobulin superfamily</td>
<td>&gt;26700</td>
<td>&gt;392</td>
<td>&gt;6</td>
<td>ND</td>
<td>75 65 (NG9)</td>
</tr>
</tbody>
</table>

Comparative Sequences of MHC Class II Region
Figure 1 (Continued on facing page)
man LINE repeats were twice as common in the classical class II region compared to ALU-SINE repeats (23% for LINEs and 9.7% for SINEs). Human SINE repeats were twice as dense in the extended class II region as their counterpart SINEs in mouse and cat (Daniels and Deininger 1985; Slattery et al. 2000). The LINE elements of human and cat are 10 times more dense in the extended class II region than LINEs in the same mouse regions. Cats had 3–6 times fewer endogenous retroviral long terminal repeat (LTR) sequences across their MHC compared to mouse or human, whereas mice showed a 4–8-fold reduction in DNA transposon elements relative to human and cat (Table 2).

The number of STR-microsatellite motifs was highest in the cat (285 loci), lowest in human (149 loci), and intermediate in the mouse (186 loci). The reduction in human was greater when loci with counts of greater than eight perfect repeats (di- and trinucleotide) were considered. The frequency of dinucleotide repeats in the human class II sequence was three times lower than in domestic cat and mouse. Trinucleotide repeats with more than eight core repeats were not found in the human sequence; there were some in cat and mouse sequence.

**DISCUSSION**

The complete MHC class II region sequence analysis allowed a direct comparison among species of three mammal orders.
Figure 2. Comparative genomic organizations of the MHC class II region in mouse, domestic cat, and human. Brackets (A–C) indicate gene segments in human, but absent in mouse and cat. (D) A human gene segment absent in cat. (E) A mouse gene segment absent in human and cat. (F) A mouse segment absent in cat. Three feline LINE1 elements and unconfirmed open reading frames (X) shown in Table 1 are not included in the three species' maps. Sequences with accession nos. AF110520, AF100956, AF027865, AF050157 for mouse and Z97183, Z97184, AL031228, AL049813, X02228, X03100, Z95437, Z81310, Z99705, Z84497, Z96104, X87344, Z84490, AL049729, Z80899, U92032, Z84489, AL049740, Z84814, AL034394 for human were used for this figure.
which last shared a common ancestor around 80 million years ago. The data showed a remarkable conservation of nucleotide sequence and gene organization, especially between human and domestic cat. This conservation was evident in the extended class II region and in portions of the classical class II region encoding DOA (DNA), RING3 kinase, DMA, B, LMP2, TAP1, LMP7, TAP2, and DOB genes. In contrast, subregions encoding classical class II antigen presenting molecules, DP, DQ, and DR were thoroughly reorganized in both species. In the human, functioning pairs of α- and β-DP, DO, and DR genes are maintained, whereas in domestic cat, only DR functional genes were evident.

One pair of DPB and DPA genes was identified in the domestic cat, but several lines of evidence suggest that these are pseudogenes. First, the cat DPA gene had higher sequence similarity (70%) and gene order syntetic homology to the human DPA2 pseudogene rather than to the functional human DPA1 gene (53% similarity). Second, the cat DPA gene had a premature termination codon resulting from a single base pair deletion in an extracellular domain coding exon. Third, a direct sequence comparison confirmed evidence that human DPA3 and cat DPA pseudogenes shared similar characteristics of insertion and deletion in the transmembrane and cytoplasmic domains, suggesting that human and cat
DPA pseudogenes descended from a common ancestral pseudogene (Verhoefen et al. 1988). Fourth, sequence homology between cat DPB and human DPB1 genes was found only in a portion of transmembrane and cytoplasmic domain coding region, and no homology was found in the extracellular domains between cat DPB and the functional human DPB1 genes (data not shown).

Previous attempts to isolate DQ cDNAs from the domestic cat were unsuccessful (Yuhki and O’Brien 1997), and DQ-like signals could not be detected by Southern blotting of cat DNA (Yuhki and O’Brien 1988). In addition, BAC/PAC contig mapping of the domestic cat class II region indicated that the interval between DOB and DRB4 genes was ∼60 kb (Beck et al. 2001). Because the interval between the human DOB and DRB1 is more than 200 kb, the DQ region is either missing or foreshortened in domestic cat MHC class II. The present sequence affirms that the cat MHC lacks the entire DQ subregion. The domestic cat is the first mammalian species to lack the entire DQ region among MHC loci characterized to date, including that of human, mouse, rat, cow, sheep, rabbit, pig, horse, and dog (Takahashi et al. 2000).

The DR subregion in the domestic cat was rearranged by four gene-duplications and one inversion event (Fig. 4). Sequence comparisons between all α- and β-genes indicated that DRB4 and DRA3 genes were the most divergent from other cat β- and α-DR genes sharing 69% (DRB4 vs. DRB1, DRB3) and 84% (DRA3 vs. DRA1, DRA2) nucleotide sequence identities including intron sequences. Both DRB4 and DRA3 genes appear to encode functional α- and β-proteins based on predicted amino acid sequences (data not shown). Two adjacent pairs of DRB-DRA genes (DRB1-DRA1) and (DRB3-DRA2) are more closely related, sharing 97%–98% nucleotide sequence identity including intron sequences, suggesting that these pairs of genes were generated by a recent gene duplication event (Fig. 4).

Takahashi et al. (2000) reviewed origins and divergence times of class II MHC gene clusters by examining α- and β-genes in placental, marsupial, and avian species. They concluded that α- and β-genes in each of DP, DO, DM, DQ, and DR class II subregions were established around the time of separation of placental mammals and marsupials, ∼160–260 million years ago (Takahashi et al. 2000; Eizirik et al. 2001). Most mammalian and marsupial species have α- and β-genes in these subregions. However, some of the gene clusters were apparently rearranged or deleted in the evolutionary process. For example, the mole rat MHC lacks the entire DR subregion (Nizetic et al. 1987) and contains multiple α- and β-genes in the DP subregion, whereas the bovine and ovine DP subregion was replaced by new clusters of DUDY (Stone and Muggli-Cockett 1990; van der Poel et al. 1990; Wright et al. 1994). The primate DP subregion contains two pairs of α- and β-genes with at least one pair of functional DP genes, whereas other mammalian DP subregions contain a nonfunctional pair of α- and β-genes (Takahashi et al. 2000).

The present study reveals that the domestic cat evolutionary lineage has experienced similar large changes in the class II region. The most dramatic change in the domestic cat class II region includes loss of functional DQ and DP genes and expansion of the DR gene family (Figs. 2, 3). Loss of the DQ and DP genes may be the reason for a remarkable immunological “tolerance” seen in cats from several perspectives, including: (1) the lack of cytotoxic antibody production (including IgE) in multiparous cats (Pollack et al. 1982; Winkler et al. 1989), (2) very inefficient antibody induction; only 13 of 59 domestic cats (22%) produced detectable levels of antibodies following allogeneic lymphocyte immunization and skin graft transplantation (Winkler et al. 1989), and (3) a remarkable tolerance for allogeneic bone marrow and tissue transplants among noninbred cats used in gene therapy protocols (Simonaro et al. 1999; Sun et al. 1999). The critical role of class II gene families in antibody production would support the notion that DP/DQ loss in domestic cats contributes to quantitative diminution of antibody recognition and induction. As a corollary, it is tempting to speculate that the expansion of feline DR genes may reflect an adaptive MHC compensation for the DP/DQ depletion.

The cat MHC class II region maintained a similar level of

![Figure 4](image-url)
Large-scale isolation of PAC DNA clones was performed using the QIAGEN Large-Construct Kit. Briefly, a single bacteria colony was inoculated into 3 mL of Superbroth containing 40 μg/mL kanamycin, and grown overnight with shaking at 37°C. Then 0.5 mL of the overnight culture was inoculated into 250 mL of Superbroth containing 40 μg/mL kanamycin, and shaken at 37°C for 3 h until A550 reached 0.15. Then 2.5 mL of 0.1M IPTG was added to the culture and incubated for 3 h at 37°C until A550 reached 1.5. Cells were harvested by centrifugation at 5000 rpm for 10 min. The bacterial pellet was resuspended in 10 mL of P1 buffer (QIAGEN Large-Construct Kit), and then 10 mL of P2 buffer was added, mixed gently, and incubated at room temperature for 5 min. Then 10 mL of chilled P3 buffer was added, mixed and incubated on ice for 10 min. After centrifugation at 8000 rpm for 15 min in 2059 Falcon tubes, the clear supernatant was collected and applied to a QIAGEN-tip 500 according to the manufacturer’s instructions. Large-scale isolation of BAC DNA was performed essentially as described above except that kanamycin was replaced with 25 μg/mL chloramphenicol, IPTG induction was omitted, and ATP-dependent exonuclease and ATP were added according to the manufacturer’s instructions.

**Methods**

**PAC/BAC Sequencing Ready Contigs**

A sequence-ready PAC/BAC contig for the entire classical class II region was described elsewhere (Beck et al. 2001). Briefly, this contig was generated by screening domestic cat PAC and BAC (RPCI 86) libraries using homologous MHC cDNA and PCR clones as probes. Representative clones for a large-scale sequencing were selected in order to minimize the amount of sequencing effort. The PAC/BAC clones 186b21, 102h1, f20, f23, 463h11, 6b1, 244e2, f2, g7, and 160a17 were selected to cover the entire extended and classical class II regions (BAC clones are indicated in bold).

**Large-Scale Isolation of PAC and BAC DNAs**

Large-scale isolation of PAC DNA clones was performed using the QIAGEN Large-Construct Kit. Briefly, a single bacteria colony was inoculated into 3 mL of Superbroth containing 40 μg/mL kanamycin, and grown overnight with shaking at 37°C. Then 0.5 mL of the overnight culture was inoculated into 250 mL of Superbroth containing 40 μg/mL kanamycin, and shaken at 37°C for 3 h until A550 reached 0.15. Then 2.5 mL of 0.1M IPTG was added to the culture and incubated for 3 h at 37°C until A550 reached 1.5. Cells were harvested by centrifugation at 5000 rpm for 10 min. The bacterial pellet was resuspended in 10 mL of P1 buffer (QIAGEN Large-Construct Kit), and then 10 mL of P2 buffer was added, mixed gently, and incubated at room temperature for 5 min. Then 10 mL of chilled P3 buffer was added, mixed and incubated on ice for 10 min. After centrifugation at 8000 rpm for 15 min in 2059 Falcon tubes, the clear supernatant was collected and applied to a QIAGEN-tip 500 according to the manufacturer’s instructions. Large-scale isolation of BAC DNA was performed essentially as described above except that kanamycin was replaced with 25 μg/mL chloramphenicol, IPTG induction was omitted, and ATP-dependent exonuclease and ATP were added according to the manufacturer’s instructions.

**Construction of Shotgun Libraries**

Shotgun sequencing libraries were constructed from purified PAC/BAC DNA. Briefly, 20 to 80 μg PAC or BAC DNA was nebulized using an Aero-Mist Nebulizer (CIS-US) for 15 to 20 min at 4 psi using argon gas in a total volume of 1.5 mL containing 35 mM Tris-HCl pH8.0, 15 mM MgCl2, and 25% glycerol. The size range of sheared DNA fragments was assessed by electrophoresis through 1% agarose gel. When the size range reached 1.5 to 3.0 Kb, the DNA was precipitated by addition of 0.1 volumes of 3 M NaOAc, pH 7.5 and 2.5 volumes of ethanol by centrifugation at 14,000 rpm for 10 min in an Eppendorf centrifuge. The DNA pellet was washed with 70% EtOH and dried briefly. End structures of randomly sheared DNA fragments were repaired by resuspending sheared DNA fragments in 25 μL of TM buffer (35 mM Tris-HCl, pH 8.0, 15 mM MgCl2), then addition of 10 μL of 5 × forward reaction buffer (Gilco BRL), 5 μL of 10 mM rATP (Pharmacia Biotech), 7 μL of 0.25 mM dNTPs (PE Applied Biosystems), 3 units T4 polyadenylate kinase (Gilco BRL) and 10 units of Klenow fragment DNA polymerase (Gilco BRL) followed by incubation at 37°C for 30 min. After fractionation through a 1% agarose gel, agarose blocks containing DNA fragments ranging in size from 1.5 to 3.0 Kb were excised and the DNA fragments were electroeluted into dialysis tubing (Gilco BRL) and then ethanol-precipitated as above. The DNA pellet was resuspended in 20 μL of TE, and a 10 μL aliquot was ligated to 100 ng of pBlueScript KS+ (Stratagene) that had been digested with EcoRV and dephosphorylated.

**Comparative Sequences of MHC Class II Region**

**Table 2.** Distribution of Repetitive Elements in Human, Cat, and Mouse MHC Regions

<table>
<thead>
<tr>
<th></th>
<th>Extended class II region</th>
<th>Classical class II region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>human</td>
<td>cat</td>
</tr>
<tr>
<td></td>
<td>278,351 bp</td>
<td>232,000 bp</td>
</tr>
<tr>
<td>No. %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINEs</td>
<td>295</td>
<td>151</td>
</tr>
<tr>
<td>LINES</td>
<td>48</td>
<td>59</td>
</tr>
<tr>
<td>LTRs</td>
<td>29</td>
<td>14</td>
</tr>
<tr>
<td>DNA transposon</td>
<td>35</td>
<td>26</td>
</tr>
<tr>
<td>STRs*</td>
<td>39</td>
<td>90</td>
</tr>
<tr>
<td>di</td>
<td>26</td>
<td>84</td>
</tr>
<tr>
<td>tri</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>tetra</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>penta</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>hexa</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*STRs with more than six perfect repeats were counted.

GC content to human HLA in both extended and classical class II regions. The cat class II region maintained 49.1% and 41.2% GC levels in the extended and classical regions, respectively, whereas the human HLA class II region maintained 49.9% and 41.3% in its same regions, sharing the same character of isochores (H2 and L2; Bernardi 2000a,b). In contrast, the mouse H2 class II region showed nearly identical levels of GC content in the extended (51.7%) and classical (47.4%) regions, both categorized as isochore H2. This observation supports the finding that the GC-poor classical class II region is significantly shorter in mouse than in human (Pavlicek et al. 2002). This may reflect the low copy number of LINE repeats and the high gene density in the mouse MHC class II region. Alternatively, a loss of GC-poor DNA in an ancestral mouse is in agreement with a general propensity of more compact vertebrate genomes to contract their DNA preferentially in the GC-poor regions (Pavlicek et al. 2002). In addition, the distinct levels of interspersed repeats, STRs, and GC levels observed in these three mammalian MHC class II sequences (Table 2) may reflect unique histories of each mammalian genome and their included major histocompatibility complex.
with calf intestinal alkaline phosphatase, using 1× T4 DNA ligase buffer and 1 unit T4 DNA ligase at 4°C for 16 h. After ethanol precipitation, the recovered DNA ligation products were resuspended in 10 µL of H2O, and a 1.0 µl aliquot was electroporated into 25 µL of E. coli DH10B cells ( Gibco BRL) at 1.6 kV. After addition of 975 µL of SOC media (Gibco BRL) and incubation for 30 min at 37°C, 50 µl aliquots were spread onto LB agar plates containing 50 µg/ml ampicillin which had been pre-treated by spreading 50 µL Xgal (25 mg/mL) and 10 µL of IPTG (100 mM) just before use. The plates were incubated overnight at 37°C.

High-Throughput Mini-Preparation and Sequencing of Plasmid DNAs

High-throughput mini-preparations were performed essentially as described by Ng et al. (1996). Bacterial colonies from the shotgun library were picked using toothpicks and inoculated into a 2-ml 96-well plate containing 1.25 mL of Superbroth and 50 µg/mL ampicillin and two glass balls (3.5 mm in diameter). The plate was incubated overnight at 37°C with shaking at 200 rpm. Bacterial cells were pelleted by centrifugation in a Sorvall RT 6000 centrifuge at 3000 rpm for 10 min at 4°C, the medium was discarded, drained onto a paper towel, and the bacterial pellet was resuspended in 100 µL of P1 buffer by vortexing. After addition of 200 µL of P2 buffer and gentle mixing, 150 µL of chilled P3 buffer was added, rapidly mixed and incubated for 10 min on ice. After centrifugation as above, 400 µL of the supernatant was transferred into wells of a 1-ml 96-well plate. Plasmid DNA was precipitated by addition of 280 µL of isopropanol and centrifugation as above for 20 min. The DNA precipitate was resuspended in 50 µL of TE buffer and further purified by Sephacryl S-500 chromatography using a 96-well SilentScreen plate (Nalge Nunc Intl.).

High-throughput nucleotide sequencing reactions were performed in 96-well plates using a total reaction volume of 10 µL containing 2.5 µL of plasmid mini-preparation (~100 to 500 ng DNA), 2 µL Dye Terminator reaction mixture, 0.5 µL of 3.2 µM T3/T7 primer, and 1 µL of 5 x reaction buffer. Sequencing reactions were performed in a GeneAmp PCR System 9600/9700 Thermocycler (PE Biosystems) using the following reaction conditions: 95°C for 5 s; then 30 cycles of 95°C for 30 sec, 50°C for 10 sec, 60°C for 4 min; and finally cooling to 4°C. Sequencing reactions were purified using Sephadex G-50 equilibrated with H2O in a 96-well filtration plate (Millipore) by centrifugation at 2200 rpm for 5 min as described above. The purified sequencing reactions (10 µL) were analyzed using an automated DNA sequencer (ABI Prism 3700 DNA Analyzer, PE Biosystems) according to the manufacturer’s instructions. To finish PAC/BAC sequencing, bridging shotgun clones between contigs were identified using the Consed program (see below). The nucleotide sequences of plasmid DNAs were deposited in GenBank under the following accession nos.: AY152827, AY152826, AY152828, AY152829, AY152825, and AY152836, for clones 186mb1, 102b1, f20, f23, 463h11, 68b1, 244e2, f2, g7, and 160a17, respectively.

DNA Sequence Analyses

Raw data generated from the automated DNA sequencer were transferred to Biolims (ABI) in SunStation c3500 at the Advanced Biomedical Computing Center (ABCC) at NCIFrederick. A Phred base-calling program (Ewing and Green 1998) was used to read all raw data. Nucleotide sequences were examined for vector and E. coli sequences, trimmed, and assembled by the Phrap program and then edited by the Consed program (Gordon et al. 1998). Automated sequence analyses were carried out using Sun c3500, DEC-alpha, SGI high-speed computers at the ABCC. These analyses included RepeatMasker (A. Smit and P. Green, unpubl.) for identification of mammalian repeat sequences, BLASTN (Altschul et al. 1990) for nucleotide sequence homology search in databases for every sequence data file and sequence contig. Genscan (Burge and Karlin 1997) for prediction of protein coding genes, and BLASTP (Gish and States 1993) for homology search of the Genscan-predicted peptide sequences for all genes in nucleotide sequences and contigs. The Dotter (Sonnhammer and Durbin 1995) dotplot program was used to compare any two large-scale nucleotide sequences. Assembly of PAC/BAC sequences was carried out based on a previous PAC/BAC contig map. Quality control was done using Autofinish program (Gordon et al. 2001) and primer extension protocol to obtain sequence with a quality value more than 30.

ACKNOWLEDGMENTS

We thank Mary Carrington, Maureen Martin, and Xiaojiang Gao for suggestions and review of the manuscript. This publication was funded in whole or in part with Federal funds from the National Cancer Institute, NIH, under contract no. N01-C0-12400.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES


WEB SITE REFERENCES

Received November 8, 2002; accepted in revised form March 5, 2003.
Comparative Genome Organization of Human, Murine, and Feline MHC Class II Region


*Genome Res.* 2003 13: 1169-1179
Access the most recent version at doi:10.1101/gr.976103

**References**
This article cites 50 articles, 19 of which can be accessed free at:
http://genome.cshlp.org/content/13/6a/1169.full.html#ref-list-1

**Creative Commons License**
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

Cold Spring Harbor Laboratory Press