Evolution of MHC Class I Loci in Marsupials: Characterization of Sequences from Koala (Phascolarctos cinereus)

Bronwyn A. Houlden,1 Warwick D. Greville,2 and William B. Sherwin
School of Biological Science, University of NSW

We demonstrate that koala (Phascolarctos cinereus) MHC class I constitutes a variable multigene family. A total of nine partial exon 2 and 3 major histocompatibility complex (MHC) class I sequences are presented, including three loci from one koala. Variation was detected by examination of sequences from a number of individuals and family groups. The koala is the second marsupial species characterized to date, and comparisons reveal approximately 80% similarity with sequences from the red-necked wallaby (Macropus rufogriseus). The latter sequences represent at least two, and probably three, different loci. Phylogenetic analysis demonstrates that all koala sequences are more related to one another than they are to any of the wallaby loci. This indicates that the koala sequences are probably not orthologous to the wallaby genes, and thus represent a new class I gene family. In addition, marsupial gene families cluster away from human gene families, supporting a different origin of MHC genes for marsupials and eutherians.

Introduction

MHC molecules play a key role in the immune system by presenting peptides to T lymphocytes, which initiates immune response cascades. There are two classes of MHC molecules, class I and class II, which differ structurally and functionally (reviewed in Bjorkman and Parham 1990). Class I molecules consist of two chains, α and β, but only α is encoded within the MHC region (reviewed in Robinson and Kindt 1989).

In most species, the MHC class I region comprises a large multigene family, consisting of homologous loci encoding classical and nonclassical MHC products, and pseudogenes (Klein 1986). Models of class I evolution invoke frequent duplication of classical loci by unequal crossing over (Hughes 1991). Loss of expression in particular duplicated genes occurs due to the accumulation of deleterious mutations, giving rise to nonclassical genes and pseudogenes, which are eventually lost (reviewed in Nei and Hughes 1991). During this evolutionary process, orthologous relationships for class I loci are commonly identifiable from species within orders (Klein 1987). Orthologous relationships are evident for human and chimpanzee, which are widely thought to have separated 5-7 MYA (Lawlor et al. 1988; Mayer et al. 1988) and for human and cotton-top tamarin, which separated 40 MYA (Watkins et al. 1991).

However, there is no example to date of a class I locus which has an identifiable orthologue in another mammalian order. Not surprisingly then, a study of the red-necked wallaby MHC class I concluded that placental and marsupial class I loci are commonly identifiable from species within orders (Klein 1987). Orthologous relationships are evident for human and chimpanzee, which are widely thought to have separated 5-7 MYA (Lawlor et al. 1988; Mayer et al. 1988) and for human and cotton-top tamarin, which separated 40 MYA (Watkins et al. 1991).

Although characterization of marsupial MHC is very limited, Southern analysis of tammar wallaby (Macropus eugenii) has demonstrated low variability at MHC class II (McKenzie and Cooper 1994). In addition, weak or nonexistent mixed lymphocyte culture (MLC) responses have been reported in the grey short-tailed opossum (Monodelphis domestica) (Fox and Rowlands 1976; Infante et al. 1991), the koala (Phascolarctos cinereus) (Wilkinson, Kotlarski, and Barton 1992), and the Virginian opossum (Didelphis virginiana) (Rowlands 1976). This suggests that marsupials may have limited MHC class II variation, unlike most eutherian species.

In contrast to the low levels of variability found in class II, class I MHC responses have been identified in both opossums and the quokka (Setonix brachyurus) (Yadav, Waring, and Stanley 1974; Rowlands 1976; Stone et al. 1978; Infante et al. 1991), and appear to be comparable to responses in most eutherian species. However, not all eutherian species demonstrate high levels of class I MHC polymorphism. Limited MHC variation has been observed in populations of cheetah (Acinonyx jubatus) (O’Brien et al. 1985; Yuhki and O’Brien 1990), beaver (Castor fiber) (Ellegren et al. 1993) and cotton-top tamarins (Saguinus oedipus) (Watkins, Hodi, and Letvin 1988; Watkins et al. 1991). The aim of this study was firstly to gain insights into the evolution of MHC in marsupials, and secondly to demonstrate the existence of MHC variation in koalas by comparing sequences from a number of individuals and family groups...

**Methods**

Nucleic Acid Extraction

DNA was prepared from four free-ranging koalas from the following localities: **K100** from Phillip Island, Victoria; **K233** from French Island, Victoria; **K153** from Stony Rises, Victoria; and **K58** from Kangaroo Island, South Australia, as described (Taylor et al. 1991). RNA was obtained from PHA-activated peripheral blood lymphocytes of koalas resident at Featherdale Wildlife Park, Doonside, New South Wales, using the guanidinium isothiocyanate method (Chomczynski and Sacchi 1987).

Restriction Fragment Length Polymorphism (RFLP) Analysis

DNA (10 μg) was digested to completion with Taq I restriction enzyme (Pharmacia). Restriction fragments were separated by electrophoresis, then denatured, neutralized, and transferred to GeneScreen Plus membrane according to the manufacturer’s recommendation. Filters were hybridized at 37°C in the koala class I probe (clone D225.6) (spanning partial exon 2 and 3 sequence) or human **HLA-B** locus probe (Sood, Pereira, and Weissman 1981), labeled with α-32P-dCTP by the random priming method (Mega-Prime; Amersham) and washed. Sizes of DNA fragments were calculated relative to bacteriophage **λ/HindIII** DNA.

Cloning and Sequencing

RT-PCR was performed with AMV reverse transcriptase in reactions containing 1 μM of the primers 5’-**TAGCCGGAATTCTACCCACGTCGGA**C-3’ (Pphi-001) and 5’-**CCAGGTATCTGCGAAGCTT**CT-CCACGCA-3’ (Pphi-002). 1.25 mM MgCl2, 0.2 mM dNTPs and Tth polymerase (Biotech Australia). cDNA was produced at 37°C for 10 min, amplified through 5 cycles of 93°C for 30 s, 37°C for 20 s, and 72°C for 2 min, followed by 35 cycles where annealing was performed at 50°C. PCR products were digested with HindIII and EcoRI (Pharmacia). cloning into **plasmid** II KS+ (Stratagene), and used to transform competent **Escherichia coli** **NM 522** (Sambrook, Fritsch, and Maniatis 1989). Double-stranded plasmid DNA was denatured and sequenced by the dideoxy chain termination method (Sanger, Niklen, and Coulson 1977; Sambrook, Fritsch, and Maniatis 1989). The Sequence Editor package (written by Matt Clegg) was used for alignment of sequences and for computation of nucleotide similarities.

Phylogenetic Analysis

Sequence data were analyzed using the Molecular Evolutionary Genetics Analysis program (MEGA) (Kumar, Tamura, and Nei 1993). Phylogenetic trees were presented drawn using the neighbor-joining method (Saitou and Nei 1987) and were based on Jukes-Cantor distances (Jukes and Cantor 1969) or the number of differences between pairwise comparisons of sequences, using MEGA. Some analyses excluded a subset of 46 codons, which aligned with those encoding the antigen recognition site (ARS) (Bjorkman et al. 1987) in human sequences, involved in class-I-peptide-T-cell interactions. The reliability of clustering patterns was tested by bootstrapping 1,000 replicates (Felsenstein 1988). Trees were also constructed using the maximum parsimony (branch and bound) method using MEGA. A compatibility matrix was generated using the reticulate.c program (available at [http://cssmr.anu.edu.au/dmm/human.html](http://cssmr.anu.edu.au/dmm/human.html)) (Jakobsen and Eastal 1996) to detect segmental transfer (gene conversion) between different MHC sequences. The significance of the compatibility pattern was determined by generating 1,000 random matrices, which consisted of informative sites in random order.

**Results**

The exon 2 and 3 region of the MHC class I molecule was amplified from koalas using primers designed to anneal to highly conserved flanking sequences. A ~450-bp amplification product generated from mRNA was identified as MHC-class-I-homologous by hybridization with a **HLA-B** class I probe in Southern analysis (data not shown). The class I DNA fragment obtained from four unrelated individuals and two family groups (Houlden, England, and Sherwin 1996) was cloned into P**BS ZZ KS**+ and 197 clones were sequenced (table 1). Six of the nine class I sequences presented (fig. 1) were obtained independently from more than one individual (table 1). The remaining three sequences were isolated from at least two different PCR reactions, which excluded the chance of reporting sequence variation due to PCR amplification artifacts, particularly those resulting from Tth polymerase infidelity (Ennis et al. 1990).

DNA sequencing revealed a high degree of similarity between koala and red-necked wallaby MHC class I (fig. 1). Nucleotide similarity of a typical sequence F123.7 was 79% for **Maru-Mhc-UB** (clone 1.3a) (fig. 1), 82% for **Mm-u-Mhc-UA** (clone 1.1a) (data not shown), 81% for **Maru-Mhc-UA** (clone 7.1b) (data

<table>
<thead>
<tr>
<th>Clone</th>
<th>KDu</th>
<th>KFo</th>
<th>KAg</th>
<th>KBb</th>
<th>KMo</th>
<th>KMa</th>
<th>KOz</th>
<th>KT1</th>
<th>KT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1237</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>A6517</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D2256</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>D5116</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D22544</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6514</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M21122</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>D2257</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M21121</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>4</td>
<td>15</td>
<td>16</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>17</td>
<td>36</td>
<td>35</td>
<td>27</td>
<td>12</td>
<td>3</td>
<td>29</td>
<td>9</td>
</tr>
</tbody>
</table>

* Phii-MHC class I clones have GenBank accession numbers U33807–U33815.

**Koala are unrelated with the following exceptions: KMo and KMa are offspring of KBb and KMo, and KT1 is offspring of KT1.**
not shown) and 68% for HLA-A*0101 (data not shown). The koala and Maru-Mhc-UB*01 red-necked wallaby sequences were completely contiguous, whereas the other two red-necked wallaby sequences were 3 bp shorter in exon 2, corresponding to amino acid 49 in the α domain (Mayer et al. 1993).

Many features characteristic of class I proteins were conserved in the koala polypeptide (fig. 2). Comparison of the 12 marsupial sequences revealed that 24 of the 65 amino acid residues analyzed were conserved in the α1 domain, and there were 47 identical amino acids in a 74-amino-acid stretch analyzed in the α2 domain (fig. 2). Highly conserved mammalian class I amino acids included the carbohydrate addition motif asparagine-X-serine in α1 at positions 89-91, and the cystine at position 104 in α2, which is involved in intrachain disulfide bonding. Other conserved amino acids included aspartic acids at positions 29 and 30, phenylalanine at 33, proline at 47 and 60, tyrosines at 62, 87 and 163, tryptophan at 63, threonine at 67 and 146, glutamine at 75 and 99, arginine at 78, leucine at 81 and 129, the tyrosine-aspartate-glycine motif at 121-123, and the tryptophan-threonine-alanine-alanine motif at 136-139 (fig. 2). Glutamic acid at positions 64 and 166 and phenylalanine at position 88 were characteristic of marsupial class I sequences (fig. 2), instead of aspartic acid and glycine and tyrosine uniformly found at those respective positions in classical HLA sequences (reviewed in Arnett and Parham 1995).

Evaluation of phylogenetic relationships was performed using the neighbor-joining method and Jukes-Cantor distances (fig. 3A), which demonstrated that marsupial and eutherian sequences (represented by HLA sequences) formed separate branches, using chicken class I sequence as the outgroup. In addition, the neighbor-joining tree and the maximum-parsimony strict consensus tree, based on the number of differences between pairwise comparisons of sequences, were similar except for some minor rearrangement of terminal branches (data not shown).

Within the marsupial branch, the koala and wallaby sequences each formed a separate cluster (fig. 3A). That is, all class I sequences from the koala are more closely related to other koala class I sequences than they are to any wallaby sequence, and vice versa. This indicates that these sequences probably do not represent orthologous genes. Although red-necked wallaby and koala are both from the order Diprotodonta, the koala sequences thus represent a new marsupial class I gene family, which has evolved since these two distantly related marsupials shared a common ancestor 48 MYA (Springer and Kirsch 1991). Phylogenetic analysis was
Fig. 2.—Alignment of deduced amino acid sequence of the α-1 and α-2 domains of class I of koala (fig. 1), wallaby (Mayer et al. 1993), human (reviewed in Amett and Parham 1995), and chicken class I (Guillemot et al. 1988). Amino acid residues are given in the one-letter international code. Identities with the consensus sequence are indicated by ...... and insertions by ...... Numbering follows Mayer et al. (1993).
Fig. 3.—Neighbor-joining trees of MHC class I sequences using Jukes-Cantor genetic distances based on all codons (A) or excluding ARS codons (B), and showing bootstrap confidence levels for internal branches after 1,000 resamplings.
the clustering pattern was significant ($P < 0.001$). Inspection revealed a 3' region of incompatibility beginning at site 267 (fig. 4), which corresponded to the third position of codon 118, located in exon 3 (fig. 1). The matrix thus supports reticulate evolution involving a portion of exon 3, but further sequencing is required to establish the 3' boundary of the exchanged segment.

The phylogenetic relationships of the two subregions identified in the matrix were reexamined separately using the neighbor-joining tree method. For codons 29–117, which comprised the majority of the informative sites, phylogenetic relationships were essentially the same as those presented in figure 3A (data not shown). For codons 118–167, the neighbor-joining tree obtained was different, with the pairs PhcI-Mhc-U$^*$D5116 and PhcI-Mhc-U$^*$D22544, and PhcI-Mhc-U$^*$D2257 and PhcI-Mhc-U$^*$D21121 now most closely related, with bootstrap values of 30% and 100% respectively (data not shown). The identification of reticulate evolution using a compatibility matrix has provided a more complex picture of evolutionary relationships between koala class I genes than was suggested by the bifurcating trees. Further elucidation of the phylogenetic relationships between these sequences is dependent on obtaining additional molecular information of class I genes from koalas.

The phylogenetic relationships between these sequences suggest that class I in koalas is composed of similar sequences which constitute a large multigene family. This is supported by Southern hybridization of DNA restriction fragments generated with TaqI enzyme from four unrelated koalas, which revealed multiple class I homologous fragments (fig. 5). One fragment (of 2.1 kb) showed reproducible length polymorphism, but this TaqI restriction site was not found in the DNA sequences of exon 2 and 3 (fig. 1). The lower limit for the number of MHC loci in a species can be estimated from the number of restriction fragments present in Southern analysis (Nizet et al. 1985; Yuki and O’Brien 1988), which suggests that there are at least eight class I loci in koalas.

In conclusion, the class I region in koalas is a multigene family comprising polymorphic genes that have approximately 90% similarity to class I genes isolated from the wallaby. However, based on phylogenetic analysis, these genes do not appear to be orthologous counterparts. Therefore, this gene family has evolved in the 48 Myr since the two species are believed to have diverged (Springer and Kirsch 1991). In addition, all the marsupial sequences cluster away from HLA sequences (fig. 3) and other placental mammals (data not shown),
Characterization of marsupial MHC class I genes was undertaken recently in the red-necked wallaby (Mayer et al. 1993), and revealed a new family of class I loci. We have described multiple class I sequences from a second marsupial species, the koala, which show 80% nucleotide similarity with the wallaby. Of the nine sequences presented, six sequences have been characterized from one koala, indicating the presence of at least three class I loci. However, in addition to these nine sequences, it is likely that at least some of the other sequences cloned represent bona fide class I sequences from koalas. The lower estimate of eight class I genes from Southern analysis is consistent with this interpretation. The finding of a large multigene family in koalas is also consistent with the number of class I genes in other mammals (Winoto, Steinmetz, and Hood 1983; Klein 1986).

Five of these six sequences have also been identified in more than one individual, and the Phci-Mhc-U*D2256 sequence was isolated from six of the nine animals studied. This finding does not necessarily indicate a paucity in polymorphism at class I loci in koalas. Common alleles have also been found in humans. Serotyping determined that HLA-A*01 and HLA-A*02 were present in 16.9% and 28.3% of Caucasians from the U.S.A. respectively, although 19 HLA-A serotypes had been defined (Imanishi et al. 1992).

Subsequently, more sensitive molecular analysis has revealed at least 50 alleles at the HLA-A locus (reviewed in Arnett and Parham 1995), and subdivided the HLA-A*01 type into HLA-A*0101 and HLA-A*0102. However, there is no difference between HLA-A*0101 and HLA-A*0102 in the portion of class I sequence compared in this study (Arnett and Parham 1995). Thus, it is possible that nucleotide sequence obtained from other regions of koala class I genes may reveal additional alleles constituting a Phci-Mhc-U*D2256 family.

Quantification of the level of polymorphism in koala class I is not possible from our data until the allelic relationships between these sequences are clarified. We were unable to assign allelic relationships with confidence, as the sequences are almost equidistant from each other in evolutionary analysis. Further sequence information is required, but even that may be incapable of determining which sequences represent allelic counterparts and which represent homologous loci. Recent evolutionary analysis of class I sequences from the horse failed to discriminate between different loci and alleles, even with complete cDNA information (Ellis, Martin, and Morrison 1995). Thus, the koala may be a second example of species where both loci and alleles have diverged within a short time period of each other.

Evolutionary divergence of loci may also have been masked by gene conversion, or loci may have arisen by gene duplication, accounting for their considerable similarities. In fact, we detected reticulate evolution among these sequences using a compatibility matrix. This mechanism has been found to be operational in many species including mouse and human (McIntyre and Seidman 1984; Madrigal et al. 1992), and its detection in class I from koalas confirms that reticulate evolution is a widespread mode of evolution of MHC sequences.

In general, orthologous relationships for class I sequences are identifiable within mammalian taxonomic orders (Nei and Hughes 1991). Although wallaby and koala both belong to the order Diprotodonta, phylogenetic analysis indicated that class I sequences from these species were not orthologous counterparts. However, the wallaby and koala are distantly related marsupials, and it is likely that the MHC has undergone extensive evolution and diverged prior to the time when these species shared a common ancestor 48 MYA (Springer and Kirsch 1991).

Marsupials and eutherians probably diverged approximately 100-150 MYA (Hope, Cooper, and Wainwright 1990) and, as previously found for red-necked wallaby sequences (Mayer et al. 1993), koala class I sequences showed no orthology to any eutherian class I sequences. Similarly, characterization of class II in red-necked wallaby revealed two new gene families, which were designated Maru-DAB and Maru-DBB (Schneider et al. 1991). This contrasts with the relationship reported for tammar wallaby partial class Ia chain sequences, which were homologous to HLA-DNA (Slade et al. 1994), and the red-necked wallaby Maru-DNA and Maru-DRA sequences, which belonged to the eutherian DNA and DRA families, respectively (Slade and Mayer 1995). Compared with that of eutherian mammals, the organization and evolution of the MHC in marsupials is poorly understood at the present time, and requires further clarification.

It is probable that a large species divergence time between wallabies and koalas precluded the identification of orthologous class I loci. Alternately, orthologous loci may exist in wallaby and koala, but may not have been isolated for technical reasons. Koala orthologues may not have been isolated from the original red-necked wallaby cDNA library purely by chance if the library was not representative or not screened exhaustively. In addition, there is a significant mismatch between the Phci-002 primer and wallaby sequences which may have precluded the isolation of koala genes orthologous to the wallaby loci.

It is likely that the sequences presented from the koala represent functional loci because they were amplified from RNA, and therefore represent transcribed loci. Secondly, these sequences have retained many conserved amino acids characteristically found in classical class I sequences (Hughes and Nei 1989b), although they may be nonclassical class I products or pseudogenes which have not yet accumulated diagnostic mutations.
tion generally parallel those observed in well-studied eutherian animals. We have determined that orthologous relationships between class I gene families are not evident between these species. Furthermore, this data supports the hypothesis that marsupial and eutherian class I loci have a different origin.

Acknowledgments

This work was supported by a fellowship to B.H. from the W. V. Scott Foundation. We wish to thank John Stark at Featherdale Wildlife Park for sample collection.

LITERATURE CITED


Guille mot, F., A. Billaud, O. Pourquié, G. Béhar, A.-M. Chaus sä, R. Zoorob, G. Kreibich, and C. Auffray. 1988 A molecular map of the chicken major histocompatibility complex: the class II β genes are closely linked to the class I genes and the nucleolar organizer. EMBO J. 7:2775–2785.


SIMON EASTEAL, reviewing editor

Accepted June 25, 1996