Evolutionary Relationships of Major Histocompatibility Complex Class I Genes in Simian Primates

Hiromi Sawai, * Yoshi Kawamoto, † Naoyuki Takahata* and Yoko Satta*†

*Department of Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, Japan and †Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Manuscript received November 5, 2003  Accepted for publication November 16, 2003

ABSTRACT

New World monkeys (NWMs) occupy a critical phylogenetic position in elucidating the evolutionary process of major histocompatibility complex (MHC) class I genes in primates. From three subfamilies of Aotinae, Cebinae, and Atelinae, the 5′-flanking regions of 18 class I genes are obtained and phylogenetically examined in terms of Alu/LINE insertion elements as well as the nucleotide substitutions. Two pairs of genes from Aotinae and Atelinae are clearly orthologous to human leukocyte antigen (HLA) -E and -F genes. Of the remaining 14 genes, 8 belong to the distinct group B, together with HLA-B and -C, to the exclusion of all other HLA class I genes. These NWM genes are classified into four groups, designated as NWM-B1, -B2, -B3, and -B4. Of these, NWM-B2 is orthologous to HLA-B/C. Also, orthologous relationships of NWM-B1, -B2, and -B3 exist among different families of Cebidae and Atelidae, which is in sharp contrast to the genus-specific gene organization within the subfamily Callitrichinae. The other six genes belong to the distinct group G. However, a clade of these NWM genes is almost equally related to HLA-A, -J, -G, and -K, and there is no evidence for their orthologous relationships to HLA-G. It is argued that class I genes in simian primates duplicated extensively in their common ancestral lineage and that subsequent evolution in descendant species has been facilitated mainly by independent loss of genes.

Table 1 lists class I homologs of human leukocyte antigen (HLA) genes of simian primates. Note that not all of these homologs are orthologous to HLA class I genes. Although there is the 1:1 orthology among humans and great apes of A locus, there is no 1:1 orthologous relationship between HLA-A and Patr-AL (Adams et al. 2001). In Old World monkeys (OWMs), there are many A-like genes in addition to the so-called Ag gene (Boyson et al. 1996, 1997). Since these A-related genes duplicated specifically in the lineage of OWMs (Boyson et al. 1996, 1997), they are orthologous to hominoid A, but they are not 1:1 orthologs to HLA-A. In more distantly related New World monkeys (NWMs), no A-like genes have been found.

For the B and C loci, the 1:1 orthology is demonstrated within hominoids. HLA-C orthologs are found in African apes and orangutans (Chen et al. 1992; Adams et al. 1999) and this is consistent with the suggestion that the duplication of HLA-B and -C occurred >40 million years ago (MYA; Kulski et al. 1997; but see Piontkivska and Nei 2003). Since this estimate predates the divergence between hominoids and OWMs (Martin 1993), it is likely that the ancestor once possessed the HLA-C ortholog. Thus, the HLA-C ortholog seems to have been lost in gibbons and OWMs. On the other hand, orangutans and OWMs experienced gene duplication of the B locus independently, so that these genes within each species as a whole are orthologous to HLA-B. Cadavid et al. (1997) found two B-like genes in spider monkeys (Ateles belzebuth) and saki monkeys.
As for the nonclassical class I loci, E and F are well conserved and their 1:1 orthologies are largely established among primates (Otting and Bontrop 1993; Knapp et al. 1998). By contrast, the orthologous relationships of G can be seen only between hominoids and OWMs (Boyson et al. 1997; Castro et al. 2000). In NWMs, G-like genes have been reported in several subfamilies. However, the orthology of these G-like genes to HLA-G has not been established. Furthermore, in the subfamily Callitrichinae (cotton-top tamarins and marmosets), G-like genes have duplicated in genus-specific manners and may function as classical class I loci; G-like genes have been reported in several subfamilies. However, the orthology of these G-like genes to HLA-G has not been established. Furthermore, in the subfamily Callitrichinae (cotton-top tamarins and marmosets), G-like genes have duplicated in genus-specific manners and may function as classical class I loci (Watkins et al. 1999; Cadavid and Watkins 1997; Cadavid et al. 1997, 1999). In other subfamilies, there are few studies about the origin and evolution of MHC class I genes (Adams and Parham 2001).

To understand the evolutionary dynamics of primate MHC genes, it is essential to establish the orthologous relationships of various MHC loci, particularly B- and G-related genes. To this end, we design PCR primers to amplify not only exons but also the 5′-flanking region of NWM class I genes. The 5′-flanking region contains a number of phylogenetically informative insertion elements. We use this information as well as the nucleotide substitutions to study the evolutionary relationships of class I genes in simian primates.

### MATERIALS AND METHODS

**HLA class I sequences**: We retrieved genomic sequences of 10 HLA loci: HLA-A (GenBank accession nos. AP000519 and AP000520), HLA-B (AP000507), HLA-C (AP000508), HLA-E (AP000514), HLA-F (AP000521), HLA-G (AP000521), HLA-H/54 (AP000520), HLA-J/59 (AP000519), HLA-K/70 (AP000520), and HLA-L/92 (AP000516). The length of the 5′- and 3′-flanking regions is ~10 kb each and that of the coding region is ~3 kb.

**Samples**: We used the owl monkey (Aotus trivirgatus), the tufted capuchin (Cebus apella), and the spider monkey (A. belzebuth) as representatives of subfamilies of Aotinae, Cebinae, and Atelinae, respectively. We also used the rhesus monkey (Macaca mulatta) because its MHC is best studied among OWMs. We purchased the rhesus monkey genomic DNA from CLONTECH (Palo Alto, CA) and prepared the NWM genomic DNA from a 2-ml blood sample of each individual kept at the Primate Research Institute by using a blood and cell culture DNA kit (QIAGEN, Chatsworth, CA).

**PCR and sequencing**: To amplify the 5′-flanking and coding sequences of class I loci, we designed PCR primers specific to certain groups of HLA loci: pAluE, 5′-GACCCTGTCTCTCTTAAAACACAGCA-3′; pAluB, 5′-AGGATCCTAATACGATATAGGTAAC-3′; and pAluG, 5′-CTCTGTATAAGCCTGAAGGAG-3′. We used each of these PCR primers in combination with an exon 2 primer (pEX2, 5′-AAGACGCTGACCTGCAGTA-3′). PCR conditions were slightly different among primer sets.
Evolution of Primate MHC Class I Genes

Figure 1.—NJ trees of 10 HLA loci based on the \( p \)-distances in (A) the 5'-flanking region (1105 bp), (B) introns (1559 bp), (C) exons 2–3 (542 bp), and (D) exons 4–8 (462 bp). For the 5'-flanking region, HLA-E is excluded. Only bootstrap values >65% are shown.

RESULTS

HLA loci and their flanking regions: In general, the topological relationships among class I genes strongly depend on the region and the length of sequences used in the phylogenetic analysis (Parham et al. 1989, 1995). We compared the 5'- and 3'-flanking sequences of the 10 HLA loci in addition to their coding sequences. However, because the alignment in the 3'-flanking region is difficult to make, particularly for HLA-E and F, we excluded the 3'-flanking region from further analysis. We analyzed the 5'-flanking region, introns, exons 2–3, and exons 4–8 separately and evaluated the reliability of the resulting phylogenetic trees in terms of the topology and the bootstrap value (BV). As expected from extensive nonsynonymous substitutions in exons 2–3 driven by balancing selection (Hughes and Nei 1988; Takahata 1995), the tree topology in these exons appears to be much affected by homoplasy and significantly different from that in other regions. In addition, the phylogenetic relationships are least reliable in terms of BVs (Figure 1C). On the other hand, the remaining three
regions show a more or less similar tree topology, and the BVs are satisfactorily high for the 5′-flanking region and introns (Figure 1). Importantly, the 5′-flanking region contains a number of nucleotide insertions and deletions (indels), as well as Alu, L1, and L2 insertion elements. It turns out that these insertion elements and indels are phylogenetically informative, so that we subsequently focus on the 5′-flanking region.

Among phylogenetically informative insertion elements (Figure 2), there are four Alus, which allow us to classify the 10 HLA loci into four groups: B (HLA-B, -C, and -L), A (HLA-A and -H), G (HLA-G, -J, -K, and -F), and E (HLA-E). Alus are primate-specific short interspersed elements, and on the basis of diagnostic sites Alus are classified into four subfamilies: FLAM or FRAM, AluJ, AluS, and AluY (Jurka and Milosavljevic 1991; Batzer and Deininger 2002). Since these subfamilies disseminated in different evolutionary times (Willard et al. 1987; Jurka and Milosavljevic 1991; Kapitonov and Jurka 1996), HLA-B, -C, and -L share an FLAM_C by which they are grouped in B. In addition to this FLAM_C, there is an AluJo that can define a single clade of group A and group G. These groups can be distinguished from each other by an AluY specific to group A and by another AluJo specific to group G. Group E or HLA-E is highly unusual in that five Alus are specifically inserted in the 5′-flanking region. The PCR primers in MATERIALS AND METHODS are designed to include these insertion elements and we use them as phylogenetic markers to identify class I genes in OWMs and NWMS.

Classification of Mamu class I genes: When the PCR primer set was applied to M. mulatta, it yielded five different 5′-flanking sequences each of 3–4 kb in length. From the comparison of insertion elements with those of HLA, these sequences in M. mulatta could be immedi-
Evolution of Primate MHC Class I Genes

Figure 3.—Five different configurations of Alu/LINEs inserted in the 5′-flanking region of Mamu class I genes. Symbols are the same as those in Figure 2.


tately classified into three groups: one into group E, one into group B, and three into group A or G (Figure 3).

We could identify the group E sequence as Mamu-E (Boyson et al. 1995) from the presence of four Alus and one L1, of which the insertion sites are identical to those of HLA-E. For the other single Mamu sequence in group B, we found that the insertion sites of one FLAM_C and three L1s are identical to those in HLA-L (Figures 2 and 3). Therefore this sequence is designated as Mamu-L. To distinguish whether the three sequences belong to group A or G, the group-A-specific AluY is not a useful marker since it is found mainly in hominoids and rarely in OWMs or NWMs (Batzer and Deininger 2002). Instead, we distinguished group G from group A by the presence or absence of the group-G-specific AluJo. In other words, if this AluJo is present, a sequence is identified as a member of group G, and if not, it is of group A. In this way, two of the three Mamu sequences are classified into group G and one into group A. The latter is designated as Mamu-A (Miller et al. 1991) and this designation is also supported by BLAST search (see below).

Furthermore, according to indels that are shared with HLA-K or -F, the two Mamu sequences in group G are identified as Mamu-K and -F. Mamu-K shares a single 95-bp insertion with HLA-K, while Mamu-F (Otting and Bontrop 1993) shares nine indels (one 2-bp, one 6-bp, and one 120-bp insertion, and two 1-bp, two 3-bp, and two 4-bp deletions) with HLA-F. Although our failure to detect Mamu-B and -G (Boyson et al. 1995, 1996) is probably owed to insufficient screening of PCR products, we did newly identify Mamu-L and -K in addition to known Mamu-A, -F, and -E. These locus assignments by insertion elements and indels are also supported by BLAST search. Using each of the 5′-flanking sequences of Mamu class I genes as a query, we obtained the corresponding HLA or Patr ortholog as the best hit. Similarly, (Boyson et al. 1995) from the presence of four Alus and one L1, of which the insertion sites are identical BLAST search of an 150-bp exon 1–2 sequence showed that the Mamu-E and -F sequences are 99% identical to those in the database, respectively, and that the Mamu-A sequence is 88% identical to the Mamu-A*11 sequence. However, the Mamu-K and -L exon sequences turn out to be >90% identical to Mamu-B*08, suggesting either some confusion in locus identification or loose linkage between the 5′-flanking region and exons 1–2.

Classification of NWM class I genes: The same PCR primer set as for M. mulatta yielded 18 NWM sequences: 7 from A. trivirgatus, 5 from C. apella, and 6 from A. belzebuth (Table 2). Since the insertion sites of group-specific Alus are identical to those in HLA genes (Figure 4), it is possible to classify the 18 sequences into three groups: E, B, and G. Two sequences are classified into group E by L1 and AluSg, which are shared with HLA-E; 8 into group B by FLAM_C, which are shared with HLA-B, -C, and -L; and the remaining 8 into group G by AluJo and L2, which are shared with HLA-G, -J, -K, and -F. The apparent lack of group A is due to either PCR primer mismatches or its true absence in the NWM genome.

The two group E sequences from A. trivirgatus and A. belzebuth (Table 2) are unambiguously designated as Ast-E and Albe-E by insertion elements (Figures 2 and 4), which is consistent with the finding of conservation of E genes in simian primates (Knapp et al. 1998). Within group B sequences, there are four different constellations (B1–B4) of Alu/LINEs. In addition, there are eight informative indels ranging from 1 to 10 bp within the
TABLE 2
Classification of HLA loci and nonhuman primate MHC genes

<table>
<thead>
<tr>
<th>Species</th>
<th>Groups</th>
<th>E</th>
<th>B</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens*</td>
<td>E</td>
<td>HLA-E</td>
<td>HLA-F</td>
<td>HLA-A</td>
<td>HLA-H</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>HLA-B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>HLA-C</td>
<td>HLA-G</td>
<td>HLA-I</td>
<td>HLA-J</td>
</tr>
<tr>
<td>M. mulatta* (Mamu)</td>
<td>E</td>
<td>Mamu-E</td>
<td>Mamu-F</td>
<td>Mamu-A</td>
<td>Mamu-K</td>
</tr>
<tr>
<td>A. trivirgatus* (Aotr)</td>
<td>E</td>
<td>Aotr-E</td>
<td>Aotr-F</td>
<td>Aotr-B1</td>
<td>Aotr-B2</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Aotr-B2</td>
<td>Aotr-G1</td>
<td>Aotr-B3</td>
<td>Aotr-B3</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Aotr-B3</td>
<td>Aotr-G1</td>
<td>Aotr-G1</td>
<td>Aotr-G2</td>
</tr>
<tr>
<td>C. apella* (Ceap)</td>
<td>B</td>
<td>Ceap-B3</td>
<td>Ceap-G1*</td>
<td>Ceap-B4</td>
<td>Ceap-B4</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>Ceap-B4</td>
<td>Ceap-G1**</td>
<td>Ceap-B4</td>
<td>Ceap-B4</td>
</tr>
</tbody>
</table>

*Classification based on Alu and LINEs in the 5′-flanking region.

The HLA sequences were retrieved from GenBank.

The MHC symbol is followed by a four-letter abbreviation system (KLEIN et al. 1990; BONTROP and KLEIN 1997).

Ceap-G1 and Ceap-G1* are presumably allelic.

1.5-kb region upstream from exon 2. Most of these consistently support the clustering between Aotr-B1 and Atbe-B1, between Aotr-B2 and Atbe-B2, and between Aotr-B3 and Ceap-B3 (Figure 4). Since these three pairs occur between different subfamilies or families, it is likely that they are orthologous. However, some exceptional indels do not support the sister relationships of Atbe-B3 and Aotr-B3. Ceap-B4 is also exceptional in that it has a number of unique indels.

Among eight group G sequences (Table 2), Atbe-F and Aotr-F differ from the rest of group G sequences in that a diagnostic AluJo splits into two parts by a 1-kb insertion sequence of unknown origin (Figure 4). These two genes are orthologous to HLA-F and Mamu-F (OTTING and BONTROP 1993), as supported also by the presence of eight indels (one 2-bp insertion and two 1-bp, two 3-bp, two 4-bp, and one 6-bp deletions), which are uniquely shared among these four sequences (Figure 4). Among the remaining six sequences in group G, there are four informative indels, all of which are shared only among Ceap-G1, Ceap-G1*, and Aotr-G1. Ceap-G1* is unique in possessing AluSc. However, since Ceap-G1 and Ceap-G1* are sampled from a single individual of C. apella and polymorphic Alus in MHC alleles are not rare (KULSKI et al. 2001), they are likely to be allelic. For the remaining three genes, Aotr-G2, Ceap-G2, and Atbe-G2, no phylogenetic informative indels can further distinguish these genes from each other.

In short, insertion elements and indels have successfully identified HLA-E and -F orthologs (WATKINS et al. 1990; KNAPP et al. 1998) as well as group B (CADAVID et al. 1997) and group G sequences in NWMs (WATKINS et al. 1990). However, this method did not unambiguously classify all the amplified sequences into loci. For this reason and to make quantitative arguments, we carried out the phylogenetic analysis of class I sequences spanning from the 5′-flanking region to exon 2.

Phylogenetic analysis of NWM class I genes: In the phylogenetic analysis of the 5′-flanking region, we excluded all group E sequences because no reliable alignment of these is feasible with other group sequences. The analysis of the remaining sequences shows that group B and G sequences form two distinct clades with 100% BVs. The common node of group B separates the L sequences from all other B-like sequences and that of group G separates the F sequences from the rest within this group (data not shown, but see Figure 5). However, the evolutionary relationships among NWM sequences within group B or G are not completely unambiguous. Therefore, we further analyzed group B and G separately to increase the number of nucleotide sites that can be aligned. Figure 5A shows the NJ tree of 12 sequences in group B and Figure 5B shows the NJ tree of 14 group G and 3 group A sequences (HLA-A, -H, and Mamu-A).

In group B (Figure 5A), the eight NWM sequences form a single cluster together with HLA-B and -C (100% BV) to the exclusion of HLA-L and Mamu-L. Apart from the HLA-B/C lineage, there are four distinct clades of the NWG sequences: B1 is [Aotr-B1, Atbe-B1], B2 is [Aotr-B2, Atbe-B2], B3 is [Aotr-B3, Ceap-B3, Atbe-B3], and B4 is [Ceap-B4]. Both B1 and B2 clades are supported by 100% BVs, while the B3 clade is supported by 83% BV. These results are consistent with our previous assumptions that Atbe-B3 is orthologous to Aotr-B3 and Ceap-B3 and that Ceap-B4 is an independent locus. Importantly, Figure 5A suggests that the sequences in the NWM B2 clade (hereafter designated as NWM-B2 for the locus) are more closely related to HLA-B/C than those in the NWM B1, B3, and B4 clades (NWM-B1, B3, and B4, respectively). It is possible that some NWM loci in group B are orthologous to HLA-B (CADAVID et al. 1997), since HLA-B and -C duplicated in the stem lineage leading to OWMs and HLA-B-related genes are more prevalent than HLA-C in primates. Although the clade of NWM-B2 and HLA-B/C is only weakly supported by BVs, it does suggest that there is an ortholog of HLA-B in NWMs. Furthermore, the NWM-B1, -B2, and -B3 genes are shared between different families of Cebidae and Atelidae.

Of 14 group G sequences (Figure 5B), Atbe-F and Aotr-F are orthologous to HLA-F and Mamu-F and the clade of these four genes is supported by 100% BV. The remaining NWG sequences form two clades: the G1 clade consisting of Aotr-G1, Ceap-G1, and Ceap-G1* (88% BV) and the G2 clade consisting of Aotr-G2, Ceap-G2, and Atbe-G2 (94% BV). The G1 and G2 clades are also
supported by deletions and are regarded as representing different loci (NWM-G1 and -G2). Like NWM-B1, -B2, and -B3, NWM-G1 and -G2 are shared among two or three different subfamilies. It should be noted that the total clade of G1 and G2 is supported by 99% BV to the exclusion of HLA-A, -G, -J, -K, and -F, suggesting that all of these loci are paralogous to each. Ceap-G1 and Ceap-G1* are more closely related to each other than to any other sequences and are clustered in a single clade (Figure 5B). The p-distance (3.2%) between Ceap-G1 and Ceap-G1* is relatively large, but it is within a range of the observed allelic diversity at MHC loci (Satta 1993).

To examine to what extent the phylogenetic tree based on exons or introns is consistent with that of the 5′-flanking region, we sequenced the entire coding regions of four genes of A. trivirgatus (Aotr-G2, Aotr-F, Aotr-B1, and Aotr-B2). These genes individually represent members of major clades in Figure 5. The phylogenetic analysis of these coding sequences with nine HLA (excluding HLA-E) loci reveals that the tree topology is almost the same as that for the 5′-flanking region. The topology based on the 5′-flanking region is also in good agreement with that based on introns (data not shown), suggesting that phylogenetic signals in the 5′-flanking region and introns are not shuffled by recombination. However, it is substantially different from the topology based on exons 2–3.

It is generally accepted that exon 4–8 sequences represent locus specificity better than the remaining coding
region (Parham et al. 1989, 1995). We compared the phylogenetic tree for exons 2–3 or exons 4–8 with that for the 5′-flanking region. Exon 4–8 sequences make HLA-B/C, Aotr-B1, and Aotr-B2 cluster with 97% BV and make HLA-F and Aotr-F cluster with 86% BV, although the evolutionary relationships within group B and G sequences cannot be resolved (data not shown). On the other hand, exon 2–3 sequences show more complicated phylogenetic relationships than do exon 4–8 sequences. In particular, group B and G sequences do not form two distinct clades. These analyses have suggested that neither exons 2–3 nor exons 4–8 contain reliable phylogenetic signals of discriminating primate class I loci.

Class I loci in the simian primate ancestor: The pairwise p-distances among four NWM paralogous B-related loci (B1–B4 in Figure 5A) range from 11.8 to 14.3%. Similarly, the distances between L (HLA-L and Mamu-L) and four NWM B-related loci range from 15.2 to 19.3%. These p-distances are still >10%, a value averaged over 20 pairs of intron sequences between humans and NWMs (O’Huigin et al. 2002), suggesting that the divergence of these loci is prior to the divergence of these species. Thus, there were at least five group B loci (four NWM B-related and L loci) before the present simian primates began to differentiate.

Likewise, the pairwise p-distances among HLA-A, -G, -J, and -K range from 11.2 to 12.0%. It is therefore likely
that these loci also had already differentiated in an early stage of primate evolution. This ancient origin of HLA-A is supported by the presence of relatively old Alu sequences in NWMs and the recent work by Piontkivska and Nei (2003) who estimated the emergence of HLA-A as recently as 35 MYA. On the other hand, the \( p \)-distances between NWM-G1 and -G2 genes range from 9.5 to 11.5%. It is not easy to judge whether they duplicated before or after the emergence of NWMs. To be conservative in the following discussion, however, we assume that NWM-G1 and -G2 loci duplicated shortly after the split of humans and NWMs. Thus, in addition to \( F \) (Figure 5B), there were at least six paralogous loci (four HLA, G1/G2, and \( F \) loci) in groups A and G in the stem lineage of simian primates.

**DISCUSSION**

Divergence times of class I loci: To date the sequence divergences within group B or G, we constructed NJ trees on the basis of the \( d \)-distances rather than on the \( p \)-distances and examined the molecular clock hypothesis by the two-cluster test (Takezaki et al. 1995). The test reveals that four group G sequences (HLA-F, Mamu-F, Aotr-F, and Atbe-F) and three group B sequences (Atbe-B2, Aotr-B2, and Ceap-B4) have evolved significantly \((P < 0.01)\) faster than other sequences within each group. However, once we removed these sequences, no rate heterogeneity remained. Thus, we compute the average height (the average \( d \)-distances from the tips to the latest common node) of groups B and G as 9.1 \( \pm 0.5 \)% and 7.3 \( \pm 0.3 \)%, respectively. If we assume that the silent nucleotide substitution rate at primate MHC loci is \( 10^{-9}/\text{site/} \text{year} \) (Satta et al. 1993), the emergence of groups B and G can be dated as \( \sim \)90 and 70 MYA, respectively. These divergences greatly predate the split between humans and NWMs, or even the emergence of prosimians (Martin 1993).

Recently, Flügge et al. (2002) and Go et al. (2003) extensively examined prosimian class I genes, and their exon sequences have revealed the monophyletic relationships with respect to simian class I genes. It would be interesting to see if Alu and LINE markers can be consistent with this observation. We found Alu, FLAM, and FRAM in the 5'-flanking region of group B, G, and E sequences. Since these Alus were dispersed in the primate genome \( \sim \)80 MYA (Batzer and Deininger 2002), it is likely that they were integrated into the prosimian genome as well. Our molecular clock analysis has also consistently provided the ancient differentiation of group B and G sequences. Therefore, identification of prosimian class I loci by this method is feasible and even legitimate to fully understand the evolution of primate class I loci.

**Oldest class I locus in primates:** In this study, we have shown that most of primate class I loci diverged from each other before the split between humans and NWMs. But which primate class I locus diverged first? Shina et al. (1999) proposed the hypothesis that the HLA region was shaped by successive duplications of a block encompassing a single HLA locus. They concluded that a block containing HLA-F is the oldest, which is also supported by the recent work by Piontkivska and Nei (2003). However, Shina et al. (1999) excluded HLA-E from their consideration, because the gene content of the E block is totally different from that of other blocks (Kulski et al. 2000). On the other hand, Piontkivska and Nei (2003) used exon 2-4 sequences in their phylogenetic analysis. Although they carefully examined the molecular clock hypothesis among various MHC genes, the phylogeny has low BVs at critical nodes and appears to be influenced by homoplasy.

On the other hand, the NJ tree of intron sequences
reveals that HLA-E diverged first when dog and pig sequences are used as outgroups (data not shown). The first divergence of HLA-E from other HLA loci is supported by insertion elements in both 5′- and 3′-flanking regions of HLA-E. In the 5′-flanking region, HLA-E does not share any insertion elements with other HLA and among them is FRAM, the oldest Alu. Similarly in the 3′-flanking region of HLA-E, no insertion elements are shared with other HLA. It is thus concluded that HLA-E diverged first and has long evolved in isolation from the others.

**Gene duplication rate in primate class I loci:** In Figure 6, we hypothesize contraction and expansion models of group B and G loci in humans and NWMs, respectively. In group B, there are one L and at least four B-related loci in the ancestral species of humans and NWMs. As discussed, the NWM-B2 locus is likely orthologous to HLA-B and this locus is retained in both humans and NWMs. However, the NWM-B1, -B3, and -B4 loci are lost in humans and HLA-B duplicated to produce HLA-C in the stem lineage of hominoids and OWMs (a diamond in Figure 5A). By contrast, all four B loci are retained but L may be lost in NWMs. In group G, there are at least five loci (HLA-G, -F, -K, -J, and NWM-G1/G2) in addition to HLA-A in the stem lineage of simian primates (Figure 5B). While the F locus is retained in both humans and NWMs, three orthologs of HLA-G, -F, and -K may become extinct in NWMs. Similarly, the ortholog of NWM-G1/G2 is lost in humans. Subsequently, there is only one duplication to generate NWM-G1 and -G2 loci (a diamond in Figure 5B). If we assume 45–55 million years of the divergence between humans and NWMs (Martin 1993; Kumar and Hedges 1998; Takahata 2001), two gene duplications generating HLA-B/C and NWM-G1/G2 took place in the descendant lineages. In addition, HLA-H duplicated from HLA-A after the divergence of human lineage from OWMs. Consequently, the duplication rate is \( \frac{3}{2} \times \frac{2}{55} = 0.027/3 \times (45 \times 2) = 0.033/\text{genome/million years} \) during this later stage of primate evolution.

For the duplication rate in the early stage of primate evolution, we note that there is no orthologous relationship between H-2 and HLA class I genes and therefore assume that there is only one common ancestral class I gene when primates diverged 80–100 MYA (Kumar and Hedges 1998; Adkins et al. 2001). The ancestral gene might have diverged into an Elike locus and an ancestral gene of group A, B, and G loci, and this ancestor may have later expanded to generate one group A, five group B, and five group G loci. Then, at least eight duplications must have taken place during the period of 25–55 million years between the divergence of primates and rodents (80–100 MYA) and the divergence of humans and NWMs (45–55 MYA). The gene duplication rate then becomes \( \frac{8}{55} = 0.15 \) to \( \frac{8}{25} = 0.32 \) per genome per million years. This rate is nearly 10 times higher than that during the later stage of primate evolution.

Thus, the gene organization was basically remodeled by extensive duplication during the early stage of primate evolution and the present repertoire has been shaped mainly by loss of loci. Such a tempo and mode of evolution in primate class I loci surely awaits further scrutiny to determine its biological significance.

This work is supported in part by a grant (no. 12304046) from the Japan Society for the Promotion of Science and in part by the Cooperation Research Program of Primate Research Institute, Kyoto University.

**LITERATURE CITED**


Boyson, J. E., S. N. McAdam, A. Gallimore, T. G. Golos, X. Liu et al., 1995 The MHC class E locus in macaques is polymorphic and is conserved between macaques and humans. Immunogenetics 41: 59–68.


Cadavid, L. F., B. E. Mejia and D. I. Watkins, 1999 MHC class I genes in a New World primate, the coton-top tamarin (Saguinus oedipus), have evolved by an active process of loci turnover. Immunogenetics 49: 196–205.


Kulski, J. K., S. Gaudieri, M. Bellgard, L. Balmer, K. Giles et al.
Kulski, J. K., P. Martinez, N. Longman

Communicating editor: S. Yokoyama