

Birth-and-Death Evolution in Primate MHC Class I Genes: Divergence Time Estimates

Helen Piontkivska and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

The major histocompatibility complex (MHC) is a multigene family that mediates the host immune response by helping T lymphocytes to recognize and respond to foreign antigens. The high degree of polymorphism and a quick turnover of the genetic loci make the evolution of MHC genes an intriguing subject of study. To understand the evolutionary pattern of this multigene family, we studied the phylogeny and divergence times of six functional MHC class I loci from primate species. On the phylogenetic trees, locus F occupies the most basal position among these loci. Our results suggest that the F locus diverged from the other MHC class I loci about 46–66 MYA. The major diversification of the other class I loci was estimated to have occurred at about 35–49 MYA, which is before the time of separation of Old World–New World monkeys. The gene duplication leading to the classical C locus in great apes appears to have occurred about 21–28 MYA. At approximately the same time the duplication of the B locus occurred in macaques. The oldest allelic lineages of A, B, and C loci in humans seem to have appeared at least 14–19, 10–15, and 13–17 MYA, respectively. Our phylogenetic analysis supports the hypothesis that the nonclassical locus F has diverged from the rest of class I loci very early in primate evolution. The overall phylogenetic pattern observed among class I genes is consistent with the model of birth-and-death evolution.

Introduction

The major histocompatibility complex (MHC) is a large multigene family that plays a key role in the adaptive immune system of vertebrates. MHC molecules are cell-surface glycoproteins that bind antigenic peptides and present them to T lymphocytes, thereby initiating the appropriate immune responses (Klein and Horejsi 1997). They are also involved in the innate immune response, interacting with natural killer cells (Adams and Parham 2001; Dixon and Stet 2001). MHC genes are classified into two groups: class I and class II genes. The α chain of class I molecules consists of three extracellular domains (α_1 , α_2 , and α_3), a transmembrane portion, and a cytoplasmic tail. Class II molecules are composed of α and β chains, and each of those chains contains two functional domains, α_1 , α_2 , and β_1 , β_2 , respectively. The so-called antigen-recognition site (ARS) is formed by certain amino acid residues of class I domains α_1 and α_2 and class II domains α_1 and β_1 (Bjorkman et al. 1987; Klein and Horejsi 1997).

Both class I and class II MHC gene families include a large number of loci and have been shown to evolve according to the birth-and-death process (Nei and Hughes 1992; Klein et al. 1993; Nei, Gu, and Sitnikova 1997). In this process new genes are created by repeated gene duplications, and some genes may later become pseudogenes or even be deleted from the genome. As a result of the birth-and-death evolution, these multigene families consist of a mixture of divergent genes, some of which have remained in the genome for a long time, and a large number of closely related genes or pseudogenes (Ota and Nei 1994; Nei, Gu, and Sitnikova 1997). A relatively slow rate of birth-and-death evolution in class II loci makes it an attractive set of genes to study the divergence times of various loci in mammals (Klein and Figueroa 1986; Hughes and Nei 1990; Takahashi, Rooney, and Nei 2000).

The longevity of these loci is relatively high; it has been estimated that most MHC class II loci originated at least 170–200 MYA (Takahashi, Rooney, and Nei 2000). In contrast, it appears that class I loci experience a much faster rate of birth-and-death evolution than class II loci (Nei and Hughes 1992). As a result, there seem to be no orthologous relationships of different class I loci among different mammalian orders (Klein and Figueroa 1986; Hughes and Nei 1989). Furthermore, the divergence of class I genes occurred so recently that even humans and New World monkeys, which diverged only about 33–35 MYA, do not share functional genes (Watkins et al. 1990a; Cadavid et al. 1997). Similarly, class I genes from two marsupial species, separated about 48 MYA, show no orthologous relationships (Houlden, Greville, and Sherwin 1996). Therefore, the turnover rate of these class I loci must be very high. However, no serious attempts have been made to infer the divergence times of these rapidly evolving loci, in part because there were not enough DNA sequence data. Recently a fair amount of sequence data has accumulated on the class I genes in higher primates, so it is now possible to evaluate the times of birth and death of several important MHC class I genes. Furthermore, analysis of genes from closely related species such as primates allows us to focus on relatively recent events of loci origin and divergence even within relatively fast-evolving multigene families. Here, we estimate the divergence time among primate MHC class I loci, using two phylogenetic approaches: the linearized tree method and the distance regression method.

Materials and Methods

Sequences Used

MHC class I gene sequences from 16 primate species were extracted from the GenBank. GenBank accession numbers of the sequences used are given in table 1. Because there is an enormous allelic variation among MHC class I genes present in human populations, only the representative sequences from the major monophyletic allelic lineages, identified by Gu and Nei (1999), were

Key words: MHC, class I, divergence time, birth-and-death evolution, primates.

E-mail: oxp108@psu.edu.

Mol. Biol. Evol. 20(4):601–609. 2003

DOI: 10.1093/molbev/msg064

© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Table 1
List of Sequences Used

Sequence Name	Accession Number
<i>Catarrhini</i> (humans, Old World monkey, and great apes)	
<i>Papio hamadryas</i> (baboon)	
Baboon.1	U35625
Baboon.2	U35624
Baboon.3	U35626
Baboon.4	U35627
Baboon.5	AF059192
Baboon.6	AF059191
<i>Pan troglodytes</i> (chimp)	
Chimp.1	M30685
Chimp.2	AF168393
<i>Pan paniscus</i> (pygmy chimp)	
Pygmy chimp.1	L39093
Pygmy chimp.2	U05578
Pygmy chimp.3	U10534
Pygmy chimp.4	U10535
Pygmy chimp.5	AF380288
Pygmy chimp.6	AF380289
<i>Macaca fascicularis</i> (Japanese macaque)	
Japanese macaque.1	U02976
Japanese macaque.2	U02977
<i>Macaca mulatta</i> (rhesus monkey)	
Rhesus macaque.1	U41837
Rhesus macaque.2	U41834
Rhesus macaque.3	U41833
Rhesus macaque.4	U41827
Rhesus macaque.5	U41835
Rhesus macaque.6	U41829
Rhesus macaque.7	U42837
Rhesus macaque.8	AF157402
Rhesus macaque.9	AF157403
Rhesus macaque.10	U41380
Rhesus macaque.11	U41379
Rhesus macaque.12	U41826
Rhesus macaque.13	U41825
Rhesus macaque.14	U41832
Rhesus macaque.15	Z21819
<i>Hylobates lar</i> (common gibbon)	
Gibbon.1	U50091
Gibbon.2	U50090
<i>Gorilla gorilla</i> (gorilla)	
Gorilla.1	X60254
Gorilla.2	X60257
Gorilla.3	X60693
<i>Homo sapiens</i> (human)	
Human A01	U07161
Human A02	Z23071
Human A23	M64742
Human A25	M32321
Human A29	X60108
Human B07	U29057
Human B14	M24040
Human B54	M77774
Human C01	D50852
Human C02	M24030
Human C03	M99390
Human C07	AJ010749
Human E0101	M20022
Human E0102	XM_165764
Human F01	NM_018950
Human G01011	NM_002127
Human G01012	AF196177
<i>Pongo pygmaeus</i> (orangutan)	
Orangutan.1	M30681
Orangutan.2	M30680
Orangutan.3	AF118896
Orangutan.4	U50084
Orangutan.5	AY034117

Table 1
Continued

Sequence Name	Accession Number
<i>Platyrrhini</i> (New World monkeys)	
<i>Saguinus fuscicollis</i> (brown-headed tamarin)	
Brown-headed tamarin.1	U59635
Brown-headed tamarin.2	U59634
<i>Saguinus oedipus</i> (cotton-top tamarin)	
Cotton-top tamarin.1	M63952
Cotton-top tamarin.2	AF004918
Cotton-top tamarin.3	AF058916
Cotton-top tamarin.4	M63954
Cotton-top tamarin.5	M63944
Cotton-top tamarin. 6	M63947
Cotton-top tamarin.7	M63951
Cotton-top tamarin.8	M63948
Cotton-top tamarin.9	M33477
Cotton-top tamarin.10	M63945
Cotton-top tamarin.11	U49331
<i>Callithrix jacchus</i> (marmoset)	
Marmoset.1	U59640
Marmoset.2	U59641
Marmoset.3	U59637
Marmoset.4	U59638
<i>Aotus trivirgatus</i> (night monkey)	
Night monkey.1	U59645
<i>Pithecia pithecia</i> (saki)	
Saki.1	U59652
Saki.2	U59655
Saki.3	U59656
Saki.4	U59654
Saki.5	U59653
Saki.6	U59651
<i>Ateles belzebuth</i> (spider monkey)	
Spider monkey.1	U59647
Spider monkey.2	U59648
<i>Saimiri sciureus</i> (squirrel monkey)	
Squirrel monkey	U59658
Rodentia	
<i>Mus musculus</i> (house mouse)	
Mouse	NM_008208

used. In particular, five monophyletic groups of HLA-A, three groups of HLA-B and four groups of HLA-C loci were represented by the following alleles: A01, A02, A23, A25, A29, B07, B14, B54, and C01, C02, C03, C07, respectively (Gu and Nei 1999). Mouse class I genes were used as an outgroup, to root the phylogeny.

Deduced amino acid sequences were aligned with the computer program Clustal X (Thompson et al. 1997). Appropriate nucleotide sequences were aligned according to the amino acid sequences alignment and visually inspected for possible errors afterwards. The alignments are available from the authors upon request.

Phylogenetic Analysis

Phylogenetic analysis was conducted using the Neighbor-Joining (NJ) tree-building method (Saitou and Nei 1987) as implemented with the computer program MEGA2 (Kumar et al. 2001). Because the extent of sequence divergence was relatively small and no strong transition-transversion bias was detected, the evolutionary distances between sequences were estimated using the Jukes-Cantor (JC) distance (Jukes and Cantor 1969). Complete nucleotide sequences of all three extracellular

domains (α_1 , α_2 , α_3) were used. All three codon positions were used. Gaps were removed from the computations using the complete-deletion option.

Two data sets were used, consisting of *Platyrrhini* (i.e., New World monkeys) and *Catarrhini* (i.e., humans, Old World monkeys, and apes) sequences. To avoid errors associated with insufficient taxon sampling (De Rijk et al. 1995; Murphy et al. 2001), human, chimpanzee, and gorilla class I sequences were used as representatives of *Catarrhini* species in the first data set, later referred to as the platyrrhine data set. Similarly, the second data set, later referred to as the catarrhine data set, includes sequences from two species of tamarins (*S. oedipus* and *S. fuscicollis*) as representatives of *Platyrrhini* clade. A total of 274 and 270 codons were used in the platyrrhine and catarrhine data sets, respectively.

The reliability of tree topologies was evaluated by the bootstrap interior branch test (Felsenstein 1985) with 500 replications, and bootstrap probability values greater than 80% were regarded as statistically significant (Sitnikova, Rzhetsky, and Nei 1995; Nei and Kumar 2000).

To examine the reliability of NJ topologies, we also constructed maximum-parsimony (MP) and maximum likelihood (ML) phylogenetic trees using the beta-version (4.0b10) of the computer program PAUP* (Swofford 2002). For each data set MP trees were generated using a heuristic search option with 10 random stepwise-addition (SA) replicates that were followed by tree bisection-reconnection (TBR) branch swapping to completion. To estimate relative branch support, bootstrap analysis (Felsenstein 1985) with 500 replicates was conducted (i.e., 500 bootstrap replications of 10SA + TBR searches). We constructed and compared 50% majority rule consensus MP trees with the NJ trees.

Maximum likelihood tree searches were conducted with the Jukes-Cantor model of nucleotide substitutions (Jukes and Cantor 1969). Reconstruction of a ML tree usually involves extensive computational efforts. However, it has been demonstrated that the most extensive search algorithm does not necessarily produces the best results (Nei, Kumar, and Takahashi 1998; Takahashi and Nei 2000) and that the combination of the bootstrap test with relatively simple search algorithms can be as efficient as more extensive searches. Here, we used a relatively simple branch swapping algorithm such as nearest-neighbor interchange (NNI) combined with the bootstrap test, rather than the extensive TBR search. The heuristic search consisted of 10 random SA followed by NNI branch swapping (i.e., 10SA + NNI). We produced 50% majority rule consensus ML trees on the basis of 100 bootstrap replications.

Divergence Time Estimation

The branch length test as implemented in the computer program LINTREE (Takezaki, Rzhetsky, and Nei 1995) was used to test the rate constancy among sequences. This test allows identifying the sequences whose evolutionary rate significantly deviates from the average rate. Following the example of Takahashi, Rooney, and Nei (2000), we used a relatively high level of significance of 0.5% to identify such deviant sequences (see asterisks

in figure 2A and B). As shown previously (Nei and Kumar 2000), even if the molecular clock assumption is violated to some extent, it is still possible to obtain reasonable time estimates. Therefore, we estimated divergence time using (1) the complete set of sequences available and (2) only those sequences that do not violate the molecular clock assumption at the 0.5% significance level. The two data sets led to similar time estimates; therefore only the results based on the complete set of sequences in each data set are presented here (see *Results*).

To account for possible rate heterogeneity across lineages, multiple calibration points (CPs) were employed for both the linearized tree method and the distance regression method, and only those CPs that provided reconcilable estimates were used in the analysis. Because not all CPs may be appropriate for a particular data set, the preference was given to CPs that provided estimates that can be best reconciled with the fossil records. The divergence time between human and orangutan was assumed to be 13 MYA, which was obtained from both the fossils record and molecular estimates (Nei and Glazko 2002). Similarly, the human-chimpanzee and *Catarrhini-Platyrrhini* divergence times were assumed to be 6 and 33 MYA, respectively (Goodman et al. 1998; Nei and Glazko 2002). Following the example of Stauffer et al. (2001), the divergence time between human and macaque lineage was set at 23.3 MYA (Harland et al. 1990). Divergence time between tamarins and marmosets was assumed to be 16 MYA (Schneider 2000). On linearized trees (see fig. 2A and B) these different CPs are shown by arrows.

The regression model considers $d_{AB} = 2rt$, where d_{AB} represents the average evolutionary distance between two gene clusters A and B, and t represents the divergence time associated with d_{AB} (Hughes and Nei 1990). The evolutionary rate was estimated to be $r = 1.86 \times 10^{-9}$ and 1.6×10^{-9} substitutions per site per million years for the platyrrhine and catarrhine data sets, respectively (JC distance for all codon positions was used).

Supplementary Material is available online through the Nei lab's databases maintained at http://mep.bio.psu.edu/databases/MHC_I/.

Results

Phylogenetic Trees of Class I Genes

A phylogenetic tree of class I genes from various *Platyrrhini* species is presented in figure 1A. In this tree human, chimpanzee, and gorilla class I sequences were used as representatives of catarrhine species. There are several noticeable features in this tree. First, human and chimpanzee genes that represent locus F constitute the most basal group on this phylogeny. This observation is consistent with the recently proposed hypothesis about the early appearance of the F locus in the primate genome. Based on the analysis of repetitive elements identified within the human MHC region, as well as other factors such as physical distance between loci and pseudogene organization (Shiina et al. 1999; Adams and Parham 2001), we concluded that the HLA-F (i.e., human MHC-F) locus or some F-related primordial locus diverged from other MHC loci rather early in primate evolution. The

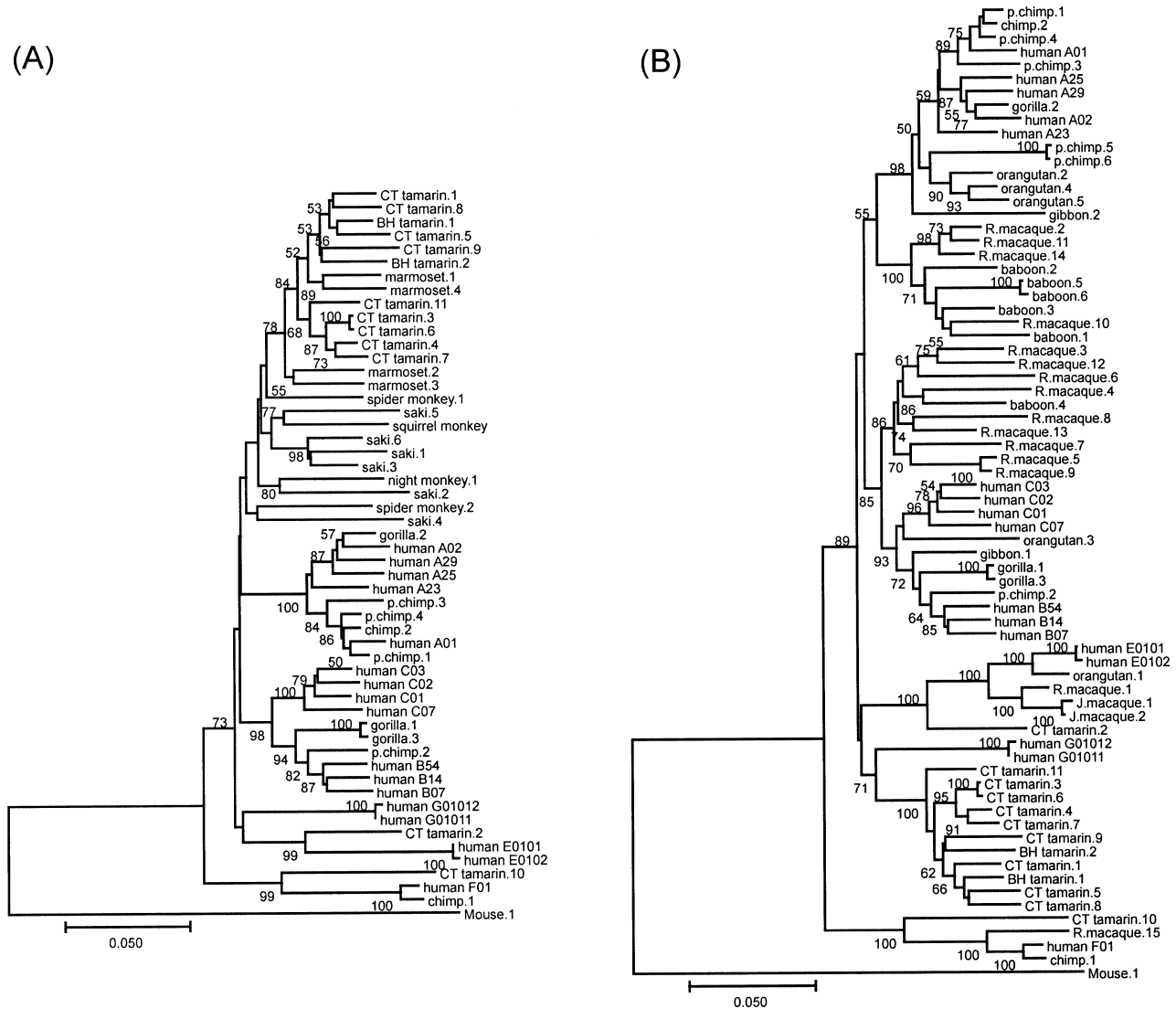


FIG. 1.—Neighbor-Joining trees for MHC class I genes for (A) platyrrhine and (B) catarrhine data sets. JC distance and all codon positions were used. Only bootstrap values above 50% are shown. Mouse sequence was used as outgroup.

presence of clearly homologous sequences in tamarins (fig. 1), as well as in several *Catarrhini* species such as chimpanzee (Lawlor et al. 1990), macaque (Otting and Bontrop 1993), and gorilla (Grimsley C., unpublished data, partial gene sequence with accession number AF159566), provides further support to the idea that locus F is a relatively ancient locus in the primate genomes. On the phylogenetic tree (fig. 1A) the appropriate sequence from tamarin (CT tamarin.10) exhibits a significant nucleotide sequence homologous to the human HLA-F gene and is clustered with the other F locus genes with high bootstrap support, suggesting orthologous relationships (Otting and Bontrop 1993; Adams and Parham 2001). Similar orthologous relationships are found among the E locus genes, where another tamarin sequence (CT tamarin.2) exhibits a high degree of homology to the human HLA-E locus genes and is clustered with them with 99% bootstrap support. Previous analysis of partial nucleotide sequences demonstrated the presence of E locus

genes in several other *Platyrrhini* species including marmosets and owl monkeys (Watkins et al. 1991; Knapp, Cadavid, and Watkins 1998). Therefore, similar to the F locus, the E locus appears to have existed in the primate genome before the *Platyrrhini-Catarrhini* split (Knapp, Cadavid, and Watkins 1998; Adams and Parham 2001).

In some species, e.g., *Aotus* and *Ateles*, the gene sequences were found to be clustered on the tree in a trans-genus-specific manner, suggesting that there are orthologous relationships between class I genes of different genera of *Cebidae*. Furthermore, in the phylogenetic tree (fig. 1A) some sequences from common marmosets (marmoset.1 and marmoset.4, in particular) were intermingled within the cluster of tamarin sequences (with 84% support for appropriate internal branch). The latter results are different from the earlier observations (Cadavid et al. 1997), where genes from different *Callitrichinae* genera, such as *Callithrix* and *Saguinus*, were found to be clustered in a genus-specific manner. However, such dif-

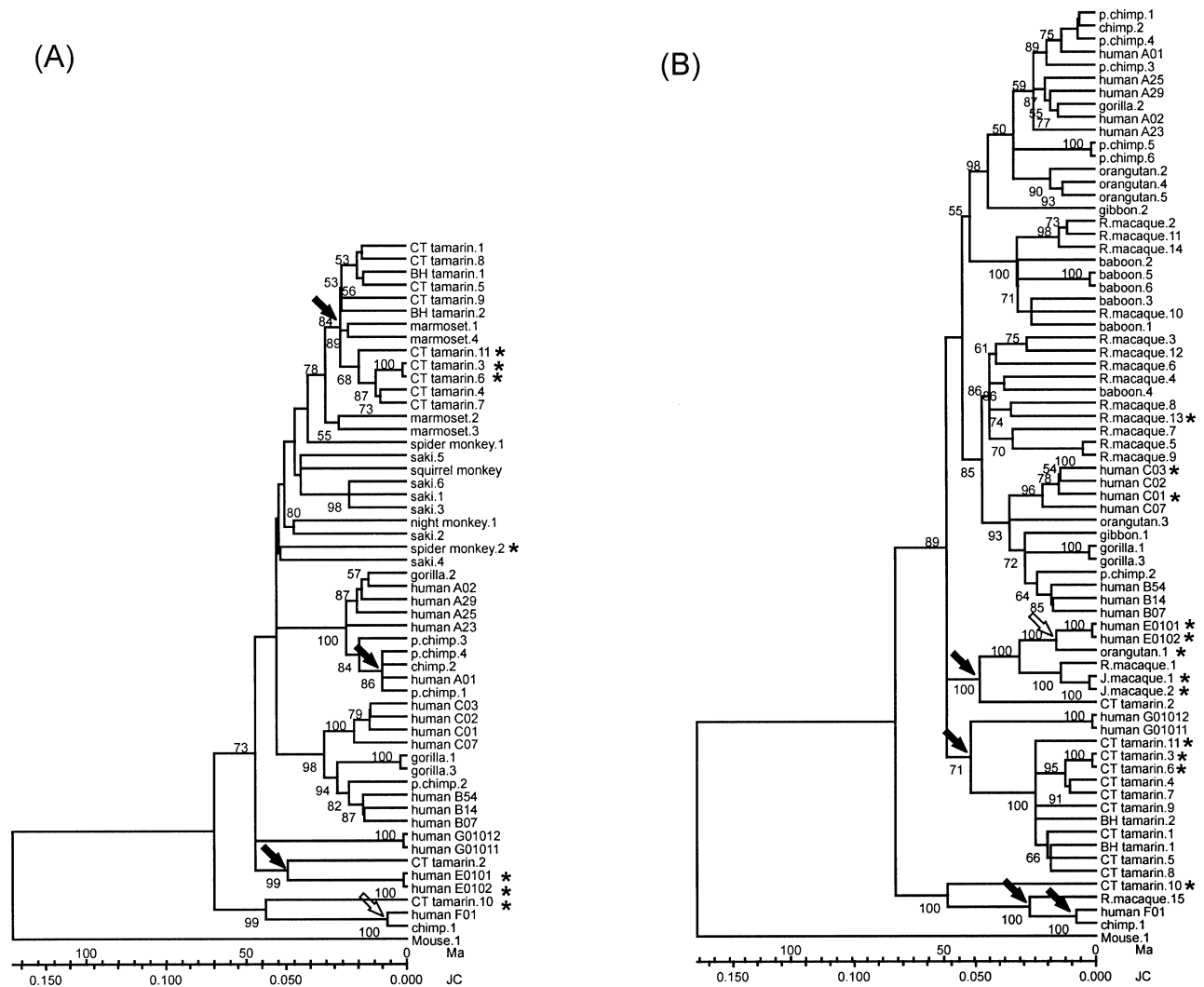


FIG. 2.—Linearized trees for (A) platyrrhine and (B) catarrhine data sets. Sequences found to be violating the molecular clock assumption at the 0.5% significance level with the branch length test are marked with asterisks (*). Here the molecular clock was calibrated with (A) 6 My divergence time between human and chimpanzee at the F locus and (B) 13 My divergence time between human and orangutan at the E locus (Nei and Glazko 2002) (CPs are indicated with a white arrow; other calibration points used are indicated with black arrows). (See *Materials and Methods* for further explanations). Only bootstrap values above 50% are shown. Mouse sequence was used as outgroup.

ferences can be attributed to the fact that only the partial sequences of exons 4 through 8 were used in the former study, whereas we used the complete sequences. The phylogenetic tree presented in figure 1A suggests that marmoset and tamarin share at least some orthologous MHC class I genes, that appear to have originated approximately at the time the *Saguinus-Callithrix* lineages split, around 16 MYA (Schneider 2000) or perhaps even earlier (see fig. 2A). However, in the absence of the complete genomic sequences from both species it is not clear whether the clustered tamarin and marmoset sequences are indeed orthologous genes. Within the genus *Saguinus* two tamarin species (the cotton-top tamarin, *S. oedipus*, and the brown-headed tamarin, *S. fuscicollis*) exhibit trans-species polymorphism at the putative G locus (Watkins et al. 1990a; Watkins 1995).

A phylogenetic tree of *Catarrhini* sequences was also constructed (fig. 1B). To avoid taxon sampling errors, sequences of two species of tamarins (*S. oedipus* and *S.*

fuscicollis) were used as representatives of *Platyrrhini* clade. As in the case of the platyrrhine data set (fig. 1A), the F locus genes of human, chimpanzee, and macaque constitute the most basal part of the phylogenetic tree. Other loci comprise separate clusters, with the classical C locus being the part of the major B locus genes cluster, confirming their common ancestry (Adams, Thomson, and Parham 1999). Furthermore, most class I genes from tamarins form a separate cluster with the human non-classical G locus (Watkins et al. 1990a; Cadavid et al. 1997) and not with the human classical loci A, B, and C. The recently described chimpanzee-specific locus AL (Adams, Cooper, and Parham 2001) is clustered with orangutan A locus sequences (sequences of pygmy chimp.5 and pygmy chimp.6, respectively), although the bootstrap support for this group is rather low. Overall, this phylogeny indicates that while the origin of Old World monkey loci, with the exception of locus C, predated the *Platyrrhini-Catarrhini* divergence, the expansion of class I

genes in New World monkey species occurred after the *Platyrrhini-Catarrhini* split.

Essentially the same tree topologies were obtained when only the first and second codon positions were used, although the bootstrap support for the majority of gene clusters decreased significantly (results not shown). Such decrease in bootstrap support can be attributed to the decrease in the number of sites involved in the resampling, arguing for the employment of as many nucleotide sites as possible (Nei and Kumar 2000).

Comparison of NJ, ML, and MP Topologies

For each data set the 50% majority rule consensus MP and ML trees were constructed (results not shown). Both the MP and ML trees showed essentially the same clustering pattern when compared with each other and with the appropriate NJ topology. The sequence clusters that received relatively high bootstrap support (80% and higher) with the NJ method were also significantly supported by the bootstrap values on the MP and ML trees. For example, the cluster confirming the common origin of B and C loci (Adams, Thomson, and Parham 1999) was supported by bootstrap values ranging from 95% to 97% on the MP trees, and 93% to 97% in ML trees. Similarly, the cluster of the F locus sequences received 99% to 100% bootstrap support in all trees examined. However, resolution of the deep divergence branches specifying the branching order of the major loci was relatively poor under the MP and ML criteria. Bootstrap support values of these branches did not exceed 50%; therefore several of these short internal branches were collapsed in 50% majority rule consensus trees. However, the lack of resolution of these branches was observed in all three methods (i.e., NJ, MP, and ML).

Because of the excessive amount of computational time, only 20 bootstrap replications of the extensive ML heuristic search (10 SA replicates followed by the TBR search) were performed. The resulting trees were then compared with the consensus trees built after 100 bootstrap replications of the less extensive search (10 SA + NNI). Overall, both heuristic procedures resulted in topologies that showed clustering patterns consistent with the appropriate NJ topologies. It appeared that use of the more extensive search, in this case TBR, does not find a better-resolved topology than the use of the simpler search algorithm, in this case NNI. Rather, both heuristic searches produced similar patterns of clustering. Furthermore, bootstrap support of the appropriate branches in NNI-based and TBR-based trees was similar, and identical to the bootstrap support of the same branches found on the NJ trees. Similar results had been observed earlier on the simulated sequence data (Nei, Kumar, and Takahashi 1998; Takahashi and Nei 2000), where use of the most extensive search algorithm cannot guarantee identification of the true tree.

Divergence Time Estimates (Linearized Trees and Distance Regression Method)

Linearized trees are presented in figure 2 (fig. 2A for platyrrhine data sets and fig. 2B for catarrhine data sets).

Using the branch length test (Takezaki, Rzhetsky, and Nei 1995), a total of 7 and 12 sequences that evolve significantly slower or faster than the average at the 0.5% level were found in the platyrrhine and catarrhine data sets, respectively (these sequences are marked with asterisks in figure 2A and B). However, estimates of divergence time for major branching points of the tree constructed using only sequences that do not violate the molecular clock assumption, were found to be quite similar to those obtained using the complete set of sequences (Nei and Kumar 2000; Takahashi, Rooney, and Nei 2000) (see also our supplementary figure 3 available online at http://mep.bio.psu.edu/databases/MHC_I/). Therefore, here we present only the results obtained using the complete set of sequences in each data set. The time scales, presented in figure 2A and B, were derived under the assumption that the human and chimpanzee F locus genes (fig. 2A) and the human and orangutan E locus genes (fig. 2B) diverged about 6 and 13 MYA (Nei and Glazko 2002), respectively. Use of other CPs, including more ancient divergences, resulted in similar time scales. Numerical results of both the linearized tree method and the distance regression method are presented in table 2. Notably, divergence time estimates obtained using these two methods are close to each other for each particular data set. Furthermore, estimates of the particular divergence times, derived from different data sets, are also very close to each other.

The divergence time estimates are presented in table 2 (see also figure 2). Our estimates showed that the F locus has existed in the primate genome for at least 46–66 My. After that, following the divergence of the F locus, at approximately 35–49 MYA, the major diversification of other MHC class I loci A, B, E, and G occurred. Therefore, these major MHC class I loci perhaps existed in the primate genome before the divergence of platyrrhine and catarrhine clades, which is believed to have occurred about 33–35 MYA (Goodman et al. 1998; Nei and Glazko 2002). Further support for this hypothesis comes from immunological assays that showed the presence of sequences homologous to classical loci A and B in some platyrrhine species, such as *Pithecia* and *Ateles* (Watkins 1995; Cadavid et al. 1997). Therefore, these loci have persisted in the primate genome for a long evolutionary time. However, other loci were found to be a result of more recent duplications. In particular, the duplication giving rise to the classical C locus in great apes occurred about 21–28 MYA. Another relatively recent duplication leading to the presence of two expressed B loci in macaque (Boyson et al. 1996) might have occurred earlier or at nearly the same time at *Cercopithecidae* lineage. The AL locus of chimpanzee was estimated to have originated 18–25 MYA (table 2). Similar estimates were obtained by Adams, Cooper, and Parham (2001), who used both coding and noncoding regions of the gene, which placed the divergence of AL and A loci at around 24 MYA.

We also estimated the age of the putatively oldest allelic lineages for human MHC class I genes. The divergence times of the major monophyletic allelic lineages (Gu and Nei 1999) were estimated as 14–19, 10–15, and 13–17 MYA for the putatively oldest alleles from classical class I loci A, B, and C in humans, respectively.

Table 2
Divergence Time Estimates Among Different Class I Loci (in Myr)

	Divergence of MHC Class I Loci in Primates							Locus B Duplication in Macaque
	F/(ABEG)	A/B/E/G	B-C Duplication	HLA-A ^a	HLA-B	HLA-C	Patr-AL ^b	
Linearized tree approach								
Platyrrhine data set	47–60	36–47	23–26	15–19	11–14	13–16	N/A ^c	N/A
Catarrhine data set	46–66	35–49	21–28	14–19	10–15	13–17	18–25	23–31
Distance regression method								
Platyrrhine data set	43.6 ± 0.6	33.0 ± 0.1	20.2 ± 1.0	12.5 ± 0.2	9.3 ± 0.8	12.0 ± 0.6	N/A	N/A
Catarrhine data set	51.3 ± 0.2	37.3 ± 0.2	23.6 ± 0.8	14.7 ± 0.2	11.5 ± 0.9	14.0 ± 0.6	21.2 ± 1.0	24.9 ± 1.5

NOTE.—Range of divergence time estimates with linearized tree approach is based on the estimates from several calibration points.

^a The age of human locus is represented as the age of the oldest allelic lineage (Klein, Sato, and O'Huigin 1998).

^b The age of the Patr-AL locus was estimated at the point of divergence of classical A loci of human and chimpanzee and A locus of orangutan.

^c N/A indicates that divergence time can not be estimated from the data set.

These results are similar to those obtained by Klein, Sato, and O'Huigin (1998), suggesting that even though the B locus is presumably older than the C locus, the divergence of its oldest allelic lineage might have occurred later than the divergence of the oldest allelic lineages of loci A and C. Furthermore, analysis by Gu and Nei (1999) showed that although the B locus is the most polymorphic of the three loci (i.e., has the highest number of alleles among all three loci), these alleles can only be separated into three groups. The largest group was comprised of several clusters; however, the bootstrap support of such allelic clusters was rather low (Gu and Nei 1999). Thus the relatively high rate of interallelic recombination, observed at this locus (Watkins et al. 1992; Marcos et al. 1997), may affect the longevity of individual alleles at the B locus, reducing the putative age of individual alleles. Alleles from human nonclassical loci E and G appeared to have diverged much later, at about 0.7–1.3 MYA (see fig. 2).

Discussion

As mentioned earlier, MHC genes have been shown to be subject to evolution by the birth-and-death process (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997). Our results are consistent with what is expected under the birth-and-death model, with class I loci experiencing frequent gene duplications and deletions, and with some genes becoming nonfunctional. An abundance of MHC class I pseudogenes, which have various degrees of sequence divergence from their functional counterparts, has been shown (Hughes 1995; Cadavid, Hughes, and Watkins 1996; Nei, Gu, and Sitnikova 1997; Boyson et al. 1996). Furthermore, while some genes apparently have stayed in the genome for a relatively long time, other loci continued to experience repeated duplications. The latter process can be illustrated by the independent expansion of G locus genes in the *Platyrrhini* and *Catarrhini* lineages after their split.

The active role that interlocus recombination might have played in shaping the evolution of MHC genes has been discussed extensively (Pease et al. 1991; Hogstrand and Bohme 1994; Yun, Melvold, and Pease 1997). In this case, the genes from several loci within species are

expected to be more similar to each other than to the genes from the other species. However, our results show that the genes from closely related species such as human and chimpanzee do not form species-specific clusters. Instead, these genes cluster in a loci-specific manner (see fig. 1). Similarly, genes from two tamarin species are intermingled on a tree, indicating that these genes have maintained their genetic identity since the speciation event. Therefore, the potential role that genetic exchanges such as gene conversion and unequal crossing-over may play in the long-term evolution of MHC genes in primates appears to be rather small. Similarly, analyzing human and mouse MHC genes, Gu and Nei (1999) showed that the overall frequency of interlocus recombination is small and that most of the genetic variation observed between MHC loci should be attributed to selection and mutation.

Our divergence time estimates suggest that some of MHC class I loci, in particular the F locus, were maintained in the primate genome for a long time, apparently at least 46–66 My. Other loci, such as A, B, G, and E, were estimated to have originated slightly later, at around 39–46 MYA, in the time period that predates the *Catarrhini-Platyrrhini* clade split (Adams, Cooper, and Parham 2001). There are also examples of more recent gene duplications leading to the appearance of new MHC class I loci. In particular, chimpanzee AL and duplicate B loci in macaque were estimated to have originated around 18–25 and 23–31 MYA, respectively (table 2), which makes them relatively young loci. Overall, our divergence time estimates are falling in agreement with the molecular estimates of the primate speciation dates (Goodman et al. 1998; Nei and Glazko 2002).

At present the question of whether some *Platyrrhini* species possess genes homologous to the classical loci of *Catarrhini* remains unclear. In particular, genes *Ateles* B*01 and *Pithecia* B*01 from spider monkey and saki, respectively, clustered significantly with the other B and C loci sequences (Watkins 1995; Cadavid et al. 1997) suggesting their orthologous relationship. In our analysis, these alleles (designated spider monkey.2 and saki.4, respectively) do not cluster with the B and C loci. Instead, they form a separate cluster within other platyrrhine sequences, although the bootstrap support value for the internal

branch was low (below 50%). Furthermore, this cluster takes a basal position among other G-locus-related *Platyrrhini* sequences, but again the bootstrap support for that pattern is rather low (see fig. 1A). Similarly, Adams and Parham (2001) observed that these two genes were not clustered phylogenetically any closer with the classical B locus than they were with the nonclassical G locus genes.

Currently, the exact order of appearance of class I loci in primates remains unclear. On the basis of the genomic structure of the MHC region in humans, it has been suggested that the duplication of the F locus led to the appearance of the G locus (Shiina et al. 1999) before other loci, such as A, B, and E, arose. The putative orthologous relationships that we observe among the locus G of great apes and humans and the appropriate loci of platyrrhine species (Watkins et al. 1990b; Cadavid et al. 1997; Adams and Parham 2001) may be used to support this idea. Our results suggest that loci A, B, and E originated rather quickly after the gene duplication that gave rise to locus G. Furthermore, our estimates place the time of origin of these major class I loci somewhat before the time of the *Catarrhini-Platyrrhini* divergence. These findings, together with the possibility that particular MHC genes have persisted in the genome for a long time, suggest that some platyrrhine species may still retain genes, orthologous to the classical A and B loci of catarrhine species. Some studies suggest that such genes may indeed exist (Watkins et al. 1990a; Watkins 1995; Cadavid et al. 1997), but the final answer to this question will come from genomic studies involving the analysis of complete genomic sequences of many primate species as well as other mammals. Current advances in complete genome sequencing and mapping should enable us to further study the evolutionary dynamics of the MHC region in a variety of organisms and to better understand the evolutionary factors affecting MHC evolution.

Acknowledgments

This work was supported by grants from the National Institutes of Health (GM20293) and the National Aeronautics and Space Administration (NCC2-1057) to M.N.

Literature Cited

- Adams, E. J., S. Cooper, and P. Parham. 2001. A novel, nonclassical MHC class I molecule specific to the common chimpanzee. *J. Immunol.* **167**:3858–3869.
- Adams, E. J., and P. Parham. 2001. Species-specific evolution of MHC class I genes in the higher primates. *Immunol. Rev.* **183**:41–64.
- Adams, E. J., G. Thomson, and P. Parham. 1999. Evidence for an HLA-C-like locus in the orangutan *Pongo pygmaeus*. *Immunogenetics* **49**:865–871.
- Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**:512–518.
- Boyson, J. E., C. Shufflebotham, L. F. Cadavid, J. A. Urvater, L. A. Knapp, A. L. Hughes, and D. I. Watkins. 1996. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *J. Immunol.* **156**:4656–4665.
- Cadavid, L. F., A. L. Hughes, and D. I. Watkins. 1996. MHC class I-processed pseudogenes in New World primates provide evidence for rapid turnover of MHC class I genes. *J. Immunol.* **157**:2403–2409.
- Cadavid, L. F., C. Shufflebotham, F. J. Ruiz, M. Yeager, A. L. Hughes, and D. I. Watkins. 1997. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. *Proc. Natl. Acad. Sci. USA* **94**:14536–14541.
- De Rijk, P., Y. Van de Peer, I. Van den Broeck, and R. De Wachter. 1995. Evolution according to large ribosomal subunit RNA. *J. Mol. Evol.* **41**:366–375.
- Dixon, B., and R. J. Stet. 2001. The relationship between major histocompatibility receptors and innate immunity in teleost fish. *Dev. Comp. Immunol.* **25**:683–699.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**:585–598.
- Gu, X., and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* **16**:147–156.
- Harland W. B., R. L. Armstrong, A. V. Cox, L. E. Craig, A. G. Smith, and D. G. Smith. 1990. A geologic time scale. Cambridge University Press, Cambridge.
- Hogstrand, K., and J. Bohme. 1994. A determination of the frequency of gene conversion in unmanipulated mouse sperm. *Proc. Natl. Acad. Sci. USA* **91**:9921–9925.
- Houlden, B. A., W. D. Greville, and W. B. Sherwin. 1996. Evolution of MHC class I loci in marsupials: characterization of sequences from koala (*Phascolarctos cinereus*). *Mol. Biol. Evol.* **13**:1119–1127.
- Hughes, A. L. 1995. Origin and evolution of HLA class I pseudogenes. *Mol. Biol. Evol.* **12**:247–258.
- Hughes, A. L., and M. Nei. 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol. Evol.* **6**:559–579.
- . 1990. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Mol. Biol. Evol.* **7**:491–514.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 in H. N. Munro, ed. *Mammalian protein metabolism*. Academic Press, New York.
- Klein, J., and F. Figueroa. 1986. Evolution of the major histocompatibility complex. *Crit. Rev. Immunol.* **6**:295–386.
- Klein, J., and V. Horejsi. 1997. *Immunology*. Blackwell Science, London.
- Klein, J., H. Ono, D. Klein, and C. O’Hugin. 1993. The accordion model of MHC evolution. *Prog. Immunol.* **8**:137–143.
- Klein, J., A. Sato, and C. O’Hugin. 1998. Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* **29**:1–21.
- Knapp, L. A., L. F. Cadavid, and D. I. Watkins. 1998. The MHC-E locus is the most well conserved of all known primate class I histocompatibility genes. *J. Immunol.* **160**:189–196.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.
- Lawlor, D. A., E. Warren, F. E. Ward, and P. Parham. 1990. Comparison of class I MHC alleles in humans and apes. *Immunol. Rev.* **113**:147–185.
- Marcos, C. Y., M. A. Fernandez-Vina, A. M. Lazaro, C. J. Nulf,

- E. H. Raimondi, and P. Stastny. 1997. Novel HLA-B35 subtypes: putative gene conversion events with donor sequences from alleles common in native Americans (HLA-B*4002 or B*4801). *Hum. Immunol.* **53**:148–155.
- Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**:614–618.
- Nei, M., and G. V. Glazko. 2002. The Wilhelmine E. Key 2001 Invitational Lecture. Estimation of divergence times for a few mammalian and several primate species. *J. Hered.* **93**:157–164.
- Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**:7799–7806.
- Nei, M., and A. L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27–38 in K. Tsuji, M. Aizawa, and T. Sasazuki, eds. 11th Histocompatibility Workshop and Conference. Oxford University Press, Oxford.
- Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.
- Nei, M., S. Kumar, and K. Takahashi. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proc. Natl. Acad. Sci. USA* **95**:12390–12397.
- Ota, T., and M. Nei. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. *Mol. Biol. Evol.* **11**:469–482.
- Otting, N., and R. E. Bontrop. 1993. Characterization of the rhesus macaque (*Macaca mulatta*) equivalent of HLA-F. *Immunogenetics* **38**:141–145.
- Pease, L. R., R. M. Horton, J. K. Pullen, and Z. L. Cai. 1991. Structure and diversity of class I antigen presenting molecules in the mouse. *Crit. Rev. Immunol.* **11**:1–32.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schneider, H. 2000. The current status of the New World monkey phylogeny. *An. Acad. Bras. Cienc.* **72**:165–172.
- Shiina, T., G. Tamiya, A. Oka et al. (23 coauthors). 1999. Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. *Proc. Natl. Acad. Sci. USA* **96**:13282–13287.
- Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**:319–333.
- Stauffer, R. L., A. Walker, O. A. Ryder, M. Lyons-Weiler, and S. B. Hedges. 2001. Human and ape molecular clocks and constraints on paleontological hypotheses. *J. Hered.* **92**:469–474.
- Swofford, D. L. 2002. *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Version 4. Sinauer Associates, Sunderland, Mass.
- Takahashi, K., and M. Nei. 2000. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.* **17**:1251–1258.
- Takahashi, K., A. P. Rooney, and M. Nei. 2000. Origins and divergence times of mammalian class II MHC gene clusters. *J. Hered.* **19**:198–204.
- Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- Watkins, D. I. 1995. The evolution of major histocompatibility class I genes in primates. *Crit. Rev. Immunol.* **15**:1–29.
- Watkins, D. I., Z. W. Chen, A. L. Hughes, M. G. Evans, T. F. Tedder, and N. L. Letvin. 1990a. Evolution of the MHC class I genes of a New World primate from ancestral homologues of human non-classical genes. *Nature* **346**:60–63.
- Watkins, D. I., T. L. Garber, Z. W. Chen, G. Toukatly, A. L. Hughes, and N. L. Letvin. 1991. Unusually limited nucleotide sequence variation of the expressed major histocompatibility complex class I genes of a New World primate species (*Saguinus oedipus*). *Immunogenetics* **33**:79–89.
- Watkins, D. I., N. L. Letvin, A. L. Hughes, and T. F. Tedder. 1990b. Molecular cloning of cDNA that encode MHC class I molecules from a New World primate (*Saguinus oedipus*). Natural selection acts at positions that may affect peptide presentation to T cells. *J. Immunol.* **144**:1136–1143.
- Watkins, D. I., S. N. McAdam, X. Liu et al. (13 coauthors). 1992. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* **357**:329–333.
- Yun, T. J., R. W. Melvold, and L. R. Pease. 1997. A complex major histocompatibility complex D locus variant generated by an unusual recombination mechanism in mice. *Proc. Natl. Acad. Sci. USA* **94**:1384–1389.

Fumio Tajima, Associate Editor

Accepted December 2, 2002