

# Gorilla Class I Major Histocompatibility Complex Alleles: Comparison to Human and Chimpanzee Class I

By David A. Lawlor, Ean Warren, Patricia Taylor,  
and Peter Parham

*From the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305*

## Summary

14 gorilla class I major histocompatibility complex (MHC) alleles have been isolated, sequenced, and compared to their counterparts in humans and chimpanzees. Gorilla homologues of HLA-A, -B, and -C were readily identified, and four Gogo-A, four Gogo-B, and five Gogo-C alleles were defined. In addition, an unusual Gogo class I gene with features in common with HLA-A and its related pseudogene, HLA-H, is described. None of the gorilla alleles is identical or even closely related to known class I alleles and each encodes a unique antigen recognition site. However, the majority of polymorphic substitutions and sequence motifs of gorilla class I alleles are shared with the human or chimpanzee systems. In particular, elements shared with HLA-A2 and HLA-B27 are found in Gogo-A and -B alleles. Diversity at the Gogo-B locus is less than at the Gogo-A locus, a trend the opposite of that seen for HLA-A and -B. The Gogo-C locus also appears to have limited polymorphism compared to Gogo-A. Two basic Gogo-C motifs were found and they segregate with distinctive sets of HLA-C alleles. HLA-A alleles are divided into five families derived from two ancient lineages. All chimpanzee A alleles derived from one of these lineages and all gorilla alleles derive from the other. Unlike chimpanzee Patr-A alleles, the Gogo-A alleles do not clearly partition with one of the HLA-A families but have similarities with two. Overall, gorilla class I diversity appears from this sampling to show more distinctions from class I HLA than found for chimpanzee class I.

The number of class I genes in the MHC varies greatly between species, as does the extent of their polymorphism (1, 2). In humans, some 17 class I genes, pseudogenes, and gene fragments have been mapped to the HLA region (3), and of these, HLA-A, -B, and -C have the widespread tissue distribution and well-developed polymorphism that characterize genes involved in presenting antigens to CD8<sup>+</sup> T cells (4). Sequence variability is focused on amino acid substitutions that change the peptide binding specificity of the antigen recognition site (5), and suggests the polymorphism results from selection for molecules that present different environmental antigens.

Comparison of alleles of the HLA-A, -B, and -C loci suggests various mutagenetic events have contributed to the accumulated polymorphism. In addition to point substitutions and intragenic recombinations (6) are recombinatorial mechanisms that move short segments of homologous sequence between alleles of the same locus and, to a lesser extent, alleles of different loci (5). Although certain alleles can easily be related in this way, others differ by highly distinctive structural motifs for which evolutionary intermediates no longer exist. Further understanding of the evolution of class I HLA

alleles may be obtained by comparison to their homologues in related species. For humans the closest relatives are the chimpanzees and the gorillas, species estimated to have all shared a common ancestor 7–10 million years ago (7–9).

Previous analysis of 10 chimpanzee class I alleles revealed a remarkable degree of similarity with HLA-A, -B, and -C alleles (10, 11). No species specific features were found and particular similarities between individual alleles in the two species were identified. Despite the similarities no example of an allele that is identical in the two species has been discovered and almost all the substitutions are at positions of the molecule predicted to alter peptide presentation. In addition, the impression obtained from what is still a limited sampling of alleles is that diversity and polymorphism in chimpanzees is but a fraction of that found in humans (12). To provide an independent assessment of these issues we now examine the class I alleles of the gorilla.

## Materials and Methods

EBV-transformed cell lines were established from Calabar, Inaki, Banga, and Oko, lowland gorillas (*Gorilla gorilla*) housed at Yerkes Regional Primate Center (Atlanta, GA). Details concerning the

individuals (age, sex, and tissue typing) and the derivation of the cell lines have been described (12). Although the ancestry of these individuals is not precisely known, there is no evidence to suggest they are related. Analysis of the class I alleles provides the following genotypes: Banga, Gogo-A\*0401, -B\*0101, -B\*0103, -C\*0101, and -C\*0201; Oko, Gogo-Ok, -B\*0201, and -C\*0202; Calabar, Gogo-A\*0101, -B\*0102, -B\*0103, -C\*0102, and -C\*0203; Inaki, Gogo-A\*0201, -A\*0501, -B\*0101, and -B\*0201. Cell lines have been deposited with the American Type Culture Collection (Rockville, MD).

The conditions for RNA extraction, first-strand cDNA synthesis, and PCR were as previously described (13). For this study, a universal 5' amplification primer, which hybridizes to all HLA class I alleles, was used. The 3' amplification primers were designed to hybridize to locus-specific regions within the 3' untranslated sequence. The reagents and procedures used by Ennis et al. (13) to target specific human class I loci were used in this study and amplification performed using the universal 5' primer in conjunction with the "locus specific" primers consisted largely of product from the target locus. Sequences of the amplification primers are: HLA-5P2, 5'-GGGCGTCGACGGACTCAGAATCTCCCA-GACGCCGAG-3'; HLA-3PALoc, 5'-CCGCAAGCTTTTGGG-GAGGGAGCACAGGTCAGCGTGGGAAG-3'; HLA-3PBloc, 5'-CCGCAAGCTTCTGGGGAGGAAACACAGGTCAGCATGGG-AAC-3'; HLA-3PCloc, 5'-CCGCAAGCTTTCGGGGAGGGAA-CACAGGTCAGTGTGGGGAC-3'.

The amplification protocol consisted of 30 cycles of denaturation (60 s at 94°C), annealing (60 s at 65°C), and extension (90 s at 72°C). Extension times were increased by 2 s with each cycle. After the final cycle, the amplification tubes were incubated at 72°C for 10 min. The amplification products were extracted and back-extracted with phenol/chloroform, ethanol precipitated, and redissolved in H<sub>2</sub>O before sequential digestions with SalI and HindIII. The digested products were ligated to similarly cut M13mp18 and mp19 before transformation of *Escherichia coli* JM109. Plaques were inoculated onto fresh cultures of JM109 in 2× yeast tryptone (YT) for overnight incubation, and single-stranded DNA was prepared from the phage lysates. Preliminary DNA sequencing was performed to group clones with identical sequence. At least three clones per group (including both orientations) were sequenced using universal and internal primers with both GTP and ITP termination mixes.

Data bases were generated with the DB program of Staden (14) or the GCG package of the University of Wisconsin (15). Alignment of the allelic sequences for exon 5 and the 3' untranslated region was determined using the multiple sequence alignment program, TULLA (16). Parsimony analysis was performed with the PAUP program (17).

## Results

**The Gorilla MHC Has Class I Genes Similar to HLA-A, -B, and -C.** 14 different class I heavy chain cDNA were cloned from four gorilla EBV-transformed B cell lines, and their nucleotide sequences were determined (Fig. 1). That locus-specific oligonucleotides based on the sequences of either HLA-A, -B, or -C alleles efficiently primed polymerase chain amplification from gorilla cDNA indicated the presence of gorilla genes related to these human class I HLA genes, and sequence comparison showed this to be the case. Alleles of the HLA-A, -B, and -C loci can be distinguished unequivocally on the basis of overall levels of sequence similarity and

the patterns of nucleotide substitution at a set of 46 "locus-specific" positions (18). Testing the 14 gorilla class I sequences by these two complementary criteria clearly shows that five segregate with HLA-A sequences, four with HLA-B sequences, and five with HLA-C sequences (Figs. 2 and 3). At the 46-locus-specific positions, two of the gorilla alleles, Gogo-A\*0201 and Gogo-A\*0401, have patterns of substitution identical to the corresponding HLA locus, and none have <37 identities (Fig. 3). Furthermore, the similarities seen in the coding region extend into the 3' untranslated sequence, a region in HLA-A, -B, and -C sequences that is divergent between loci but relatively homogenous within the alleles of a locus (Fig. 4). We find therefore that the gorilla, like the chimpanzee, has clearly recognizable HLA-A, -B, and -C-like class I MHC loci. These results demonstrate that the origin of this basic arrangement of expressed class I genes predates the divergence of humans, chimpanzees, and gorillas.

**Structural Divergence in HLA-A-like Genes of the Gorilla.** When the nucleotide sequences of Patr-A and HLA-A alleles are compared, the chimpanzee sequences give a range of pairwise differences (16–57 nucleotides) within that defined by pairwise comparisons among HLA-A alleles (2–62 nucleotides). This is true for some of the gorilla A-like sequences but not all. In particular, the Gogo-Ok sequence greatly extends the range of differences in pairwise comparisons and inspection shows this sequence has unusual features.

Closely linked to HLA-A are at least two structurally related pseudogenes with sequence motifs similar to HLA-A at the locus-specific positions and in the 3' untranslated region. Best characterized of these pseudogenes is HLA-H, which is postulated to have been an antigen-presenting locus that was inactivated through the acquisition of one or two single base pair mutations (19). Of interest here is that the Gogo-Ok sequence has similarities with HLA-H. Most striking is a 210-bp segment of exon 2 (positions 74–286) that is nearly identical in Gogo-Ok and HLA-H (Fig. 1). At the 3' end of the coding region, there are additional similarities including a termination codon distinct from those found in HLA-A and other class I HLA genes.

The patterns of substitution at locus-specific positions indicate that Gogo-Ok is equally related to both HLA-A and HLA-H and a similar conclusion is reached from overall comparisons of the sequence. For example, the range of differences between Gogo-Ok and HLA-A alleles is 62–94 nucleotides compared to 76–81 nucleotides for HLA-H alleles. Comparison of individual exons shows that Gogo-Ok is more similar to HLA-A (and other Gogo-A) alleles in exon 3 and to HLA-H alleles in exon 2 (Table 1). It is highly unlikely that the unusual pattern of sequence motifs in Gogo-Ok is the result of a PCR or cloning artifact, as an almost identical allele has been isolated by Watkins et al. (20) from another individual gorilla. These data suggest three possibilities for the nature of Gogo-Ok: first, that it is an allele of the Gogo-A locus that has been the recipient of a gene conversion(s) from an HLA-H-like gene; second, that it is an allele of an HLA-H-like gene that has been the recipient of a gene conversion(s) from a Gogo-A-like gene; and third, that

Downloaded from <https://rupress.org/jem/article-pdf/174/6/1491/492155/1491.pdf> by guest on 12 March 2020

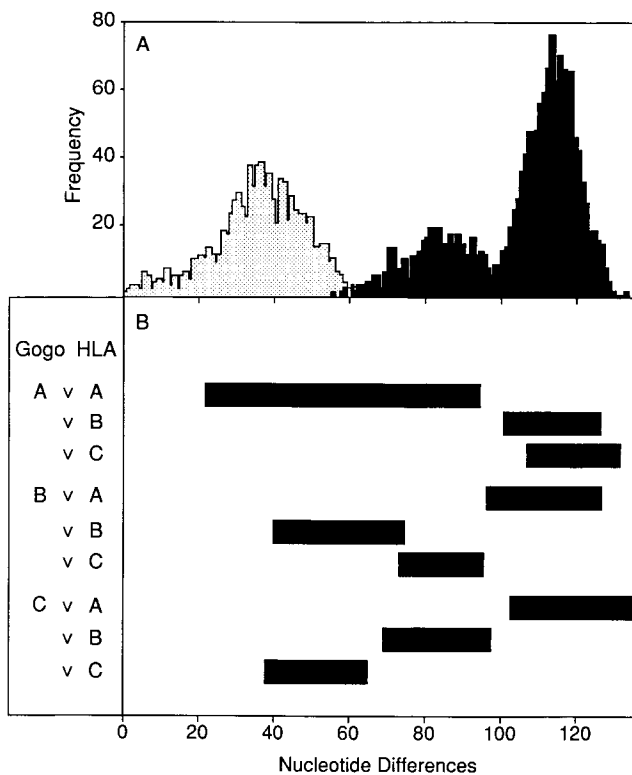
1493 Lawlor et al.

**B**

	401							480
Gogo-A*0101	TCCGCGGGTA	CCAGCAGGAC	GCCTACGACG	GCAAGGATTA	CATCGCCTTG	AACGAGGACC	TGCGCTCCTG	GACCGCGGCG
Gogo-A*0201	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0401	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-Oko	-----	-t-----	-t-----	-----	-a-c-----	-----	-t-----	-----
HLA-H	-----	tg-a-----	-----	-----	-tc-----	-----	-----	-----
Gogo-B*0101	-----	tagc-----	-----	-----	-c-----	-----	-a-----	-----
Gogo-B*0102	-----	tagc-----	-----	-----	-c-----	-----	-a-----	-----
Gogo-B*0103	-----	tagc-----	-----	-----	-c-----	-----	-a-----	-----
Gogo-B*0201	-----	tagc-----	-----	-----	-c-----	-----	-a-----	-----
Patr-B*01	-----	tggc-----	-----	-----	-c-----	-----	-a-----	-----
Gogo-C*0101	aa-----	tg-c-----	-----	-----	-c-----	-----	-----	-c-----
Gogo-C*0102	aa-----	tg-c-----	-----	-----	-c-----	-----	-----	-c-----
Gogo-C*0201	-----	tagc-----	-----	-----	-c-----	-----	-----	-c-----
Gogo-C*0202	-----	tagc-----	-----	-----	-c-----	-----	-----	-c-----
Gogo-C*0203	-----	tagc-----	-----	-----	-c-----	-----	-----	-c-----
	481							560
Gogo-A*0101	GACATGGCAG	CTGAGATTAC	CAAGCGCAAG	TGGGAGGCGG	CCCATTTTGC	GGAGCAGTTG	AGAGCCTACC	TGGAGGGCAC
Gogo-A*0201	-----	-c-----	-c-----	-----	-gag-----	-----	-----	ga
Gogo-A*0401	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-Oko	-----	-g-----	-c-----	-----	-g-gag-----	-g-----	-a-----	-----
HLA-H	-----	-c-----	-c-----	-----	-g-cgg-----	-cg-----	-t-----	ga
Gogo-B*0101	-----	cc-g-----	-c-----	-----	-g-gag-----	-c-----	-----	-----
Gogo-B*0102	-----	cc-g-----	-c-----	-----	-g-gag-----	-c-----	-----	-----
Gogo-B*0103	-----	cc-g-----	-c-----	-----	-g-gag-----	-c-----	-----	-----
Gogo-B*0201	-----	cc-g-----	-c-----	ac-----	-g-gcg-----	-ga-----	-----	-ct
Patr-B*01	-----	cc-g-----	-c-----	-c-----	-g-gag-----	-cg-----	-----	-----
Gogo-C*0101	-----	-c-g-----	-c-----	-----	-g-gg-----	-g-ca-----	-----	-ct
Gogo-C*0102	-----	-c-g-----	-c-----	-----	-g-gg-----	-g-ca-----	-----	-ct
Gogo-C*0201	-----	-c-g-----	-c-g-----	ac-----	-g-g-g-----	-g-ga-----	-----	-ct
Gogo-C*0202	-----	-c-g-----	-c-----	-t-----	-g-gcg-----	-ca-----	-----	-ct
Gogo-C*0203	-----	-c-g-----	-c-g-----	ac-----	-g-g-g-----	-ga-----	-----	-ct
	561						lex4	640
Gogo-A*0101	GTGCGTGGAG	TGGCTCCGCA	GACACTTGA	GAACGGGAAG	GAGACGCTGC	AGCGCACGGA	CGCCCCCAAG	ACGCATATGA
Gogo-A*0201	-----	-----	-t-----	-----	-----	-----	-----	-----
Gogo-A*0401	-----	-----	-t-----	-----	-----	-t-----	-----	-----
Gogo-A*0501	-----	-----	-t-----	-----	-----	-----	-----	-a-----
Gogo-Oko	-----	-----	-----	-----	-----	-----	-c-----	-a-----
HLA-H	-----	-t-----	-t-----	-----	-----	-g-----	-c-----	-a-----
Gogo-B*0101	-----	-----	-t-----	ga-----	-----	-g-----	-a-----	-a-----
Gogo-B*0102	-----	-----	-t-----	ga-----	-----	-g-----	-a-----	-a-----
Gogo-B*0103	-----	-----	-t-----	ga-----	-----	-g-----	-a-----	-a-----
Gogo-B*0201	-----	-----	-t-----	ga-----	-----	-g-----	-c-----	-a-----
Patr-B*01	-----	-----	-t-----	-----	-----	-g-----	-c-----	-a-----
Gogo-C*0101	-----	-----	-t-----	-t-----	-----	-g-a-----	-c-----	-a-----
Gogo-C*0102	-----	-----	-t-----	-t-----	-----	-g-a-----	-c-----	-a-----
Gogo-C*0201	-----	-----	-t-----	-----	-----	-g-----	ac-----	-a-----
Gogo-C*0202	-----	-c-----	-t-----	-----	-----	-g-----	ac-----	-a-----
Gogo-C*0203	-----	-----	-t-----	-----	-----	-g-----	ac-----	-a-----
	641							720
Gogo-A*0101	CTCACCACGC	TGCTCTGAC	CATGAGGCCA	TCCTGAGGTG	CTGGGCCCTG	AGCTTCTACC	CTGCGGAGAT	CACACTGACC
Gogo-A*0201	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0401	-----	-c-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	-----	-a-----	-----	-----	-----	-----	-----	-----
Gogo-Oko	-----	-c-----	-----	c-----	-a-----	-----	-----	-----
HLA-H	-----	-c-----	ca-----	-----	-----	g-----	-----	-----
Gogo-B*0101	-----	-c-----	ga-----	-----	-----	g-----	-----	-----
Gogo-B*0102	-----	-c-----	ga-----	-----	-----	g-----	-----	-----
Gogo-B*0103	-----	-c-----	ga-----	-----	-----	g-----	-----	-----
Gogo-B*0201	-----	-c-----	ca-----	-----	-----	g-----	-----	-----
Patr-B*01	-----	-c-----	ca-----	-----	-----	g-----	-----	-----
Gogo-C*0101	-----	-c-----	ca-----	-----	-----	g-----	-----	-----
Gogo-C*0102	-----	-c-----	ca-----	-----	-----	g-----	-----	-----
Gogo-C*0201	-----	-c-----	cc-----	-----	-----	g-----	-----	-----
Gogo-C*0202	-----	-c-----	cc-----	-----	-----	g-----	-----	-----
Gogo-C*0203	-----	-c-----	cc-----	-----	-----	g-----	-----	-----
	721							800
Gogo-A*0101	TGGCAGCGGG	ATGGGGAGGA	CCAGACCCAG	GACACGGAGC	TCGTGGAGAC	CAGGCCTGCA	GGGGATGGAA	CCTTCCAGAA
Gogo-A*0201	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0401	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	-----	-----	-----	-----	-----	-----	-a-----	-----
Gogo-Oko	-----	-----	-a-----	-----	-----	-g-----	-g-----	-----
HLA-H	-----	-----	-----	-c-----	-----	-----	-----	-----
Gogo-B*0101	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Gogo-B*0102	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Gogo-B*0103	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Gogo-B*0201	-----	-a-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Patr-B*01	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-a-----	-a-a-----
Gogo-C*0101	-----	-----	-c-----	-a-t-----	-tc-----	-t-----	-a-----	-a-----
Gogo-C*0102	-----	-----	-c-----	-a-t-----	-tc-----	-t-----	-a-----	-a-----
Gogo-C*0201	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Gogo-C*0202	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----
Gogo-C*0203	-----	-----	-c-----	-a-t-----	-c-----	-t-----	-a-----	-a-----

	801							880
Gogo-A*0101	GTGGGCGGCT	GTGGTGGTGC	CTTCTGGACA	GGAGCAGAGA	TACACCTGCC	ATGTGCAGCA	TGAGGGTCTG	CCCCGAGCCCC
Gogo-A*0201	-----	-----	-----	-----	-----	-----	a-----	-----
Gogo-A*0401	-----	-----	-----	-----	-----	-----	a-----	-----
Gogo-A*0501	-----	-----	-----	-----	-----	-----	a-----	-----
Gogo-Okko	-----	-----	a-----	-----	-----	-----	-----	a-----
HLA-H	-----	-----	g-----	-----	-----	-----	-----	a-----
Gogo-B*0101	-----	a-----	-----	a-----	-----	-----	g-----	ga-----
Gogo-B*0102	-----	a-----	-----	a-----	-----	-----	g-----	ga-----
Gogo-B*0103	-----	a-----	-----	a-----	-----	-----	g-----	ga-----
Gogo-B*0201	-----	a-----	-----	a-----	-----	g-----	c-----	g-----
Patr-B*01	-----	a-----	-----	a-----	-----	-----	g-----	ga-----
Gogo-C*0101	-----	a-----	-----	a-----	-----	-----	ca-----	tg-----
Gogo-C*0102	-----	a-----	-----	a-----	-----	-----	ca-----	tg-----
Gogo-C*0201	-----	a-----	-----	a-----	-----	a-----	c-----	a-----
Gogo-C*0202	-----	a-----	-----	a-----	-----	a-----	c-----	a-----
Gogo-C*0203	-----	a-----	-----	a-----	-----	a-----	c-----	a-----
	881	ex5						960
Gogo-A*0101	TCACCCTGAG	ATGGGAGCCA	TCTTCCCAGC	CCACCATCCC	CATCGTGGGC	ATCATTGCTG	GCCTGGTTCT	CTTTGGAGC.
Gogo-A*0201	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0401	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-Okko	-----	-----	g-----	-----	-----	-----	-----	tc-----
HLA-H	-----	-----	-----	-----	-----	g-----	-----	ac-----
Gogo-B*0101	-----	-----	-----	-----	-----	tg-----	-----	c-a-c-t.
Gogo-B*0102	-----	-----	-----	-----	-----	tg-----	-----	c-a-c-t.
Gogo-B*0103	-----	-----	-----	-----	-----	tg-----	-----	c-a-c-t.
Gogo-B*0201	-t-----	-----	-----	-----	-----	tg-----	-----	c-tg-tc
Patr-B*01	-----	-----	-----	-----	-----	tg-----	-----	c-tg-tc
Gogo-C*0101	-----	-----	g-----	-----	-----	tg-----	-----	c-a-c-t.
Gogo-C*0102	-----	-----	g-----	-----	-----	g-----	-----	c-a-c-t.
Gogo-C*0201	-----	-----	-----	-----	-----	tg-----	-----	c-g-tt-tc
Gogo-C*0202	-----	-----	-----	-----	-----	tg-----	-----	c-g-tt-tc
Gogo-C*0203	-----	-----	-----	-----	-----	tg-----	-----	c-g-tt-tc
	961					ex6		1040
Gogo-A*0101	.T.G.TGATC	GCT.GGA.GC	TGTGGTCGCT	GCTGTGAGGT	GGAGGAGGAA	GAGCTCAGAT	AGAAAAGGAG	GGAGCTACTC
Gogo-A*0201	.-.-.-----	-.-.-.-----	-----a-----	-----	-----	-----	-----	-----
Gogo-A*0401	.-.-.-----	-.-.-.-----	-----	-----	-----	-----	-----	-----
Gogo-A*0501	.-.-.-----	-.-.-.-----	-----	-----	-----	-----	-----	-----
Gogo-Okko	.-.-.-----	a-.-.-----	-----	a-----	a-----	-----	-----	-----
HLA-H	.-.-.g-----	a-.-.-----	-----	a-----	a-----	-----	-----	-----
Gogo-B*0101	.-.-.g-----	a-c-----	-----	a-----	a-----	g-----	-----	-----
Gogo-B*0102	.-.-.g-----	a-c-----	-----	a-----	a-----	g-----	-----	-----
Gogo-B*0103	.-.-.g-----	a-c-----	-----	a-----	a-----	g-----	-----	-----
Gogo-B*0201	acc-.a.g	tg-.--tc-	-----	-----	-----	g-----	-----	-----
Patr-B*01	acc-.a.g	tg-.--tc-	-----	-----	-----	g-----	-----	-----
Gogo-C*0101	.-.-.t-----	a-.-.a-----	-----	t-----	t-----	g-----	-----	g-----
Gogo-C*0102	.-.-.t-----	a-.-.a-----	-----	t-----	t-----	g-----	-----	g-----
Gogo-C*0201	c-a-c-----	-.t-----	-----	a-c-----	a-----	g-----	-----	g-----
Gogo-C*0202	c-a-c-----	-.t-----	-----	a-c-----	a-----	g-----	-----	g-----
Gogo-C*0203	c-a-c-----	-.t-----	-----	a-c-----	a-----	g-----	-----	g-----
	1041	ex7				ex8	1104	
Gogo-A*0101	TCAGGCTGCA	AGCAGTGACA	GTGCCCAGGG	CTCTGATGTG	TCTCTCACAG	CTTGTAAAGT	GTGA	
Gogo-A*0201	-----	-----	-----	-----	-----	-----	-----	
Gogo-A*0401	-----	-----	-----	-----	-----	-----	-----	
Gogo-A*0501	-----	-----	-----	-----	-----	-----	-----	
Gogo-Okko	-----	-----	-----	-----	-----	g-----	-----	
HLA-H	-----	g-ca-----	-----	-----	-----	g-----	-----	
Gogo-B*0101	-----	g t-c-----	-----	-----	-----	-----	a-----	
Gogo-B*0102	-----	g t-c-----	-----	-----	-----	-----	a-----	
Gogo-B*0103	-----	g t-c-----	-----	-----	-----	-----	a-----	
Gogo-B*0201	-----	g t-c-----	-----	-----	-----	-----	a-----	
Patr-B*01	-----	g t-c-----	-----	-----	-----	-----	a-----	
Gogo-C*0101	-----	g t-----	-----	a-----	-----	tc-----	c-----	
Gogo-C*0102	-----	g t-----	-----	a-----	-----	tc-----	c-----	
Gogo-C*0201	-----	g t-----	-----	a-----	-----	tc-----	c-----	
Gogo-C*0202	-----	g t-----	-----	a-----	-----	tc-----	c-----	
Gogo-C*0203	-----	g t-----	-----	a-----	-----	tc-----	c-----	

**Figure 1.** The nucleotide sequences of the coding region of 14 gorilla class I alleles. Sequences are compared to Gogo-A\*0101. Dashes signify identity and dots denote gaps introduced for optimal alignment. In bold are substitutions from the consensus in the regions of identity between Gogo-Okko and HLA-H, the polymorphism in exon 5 that is shared by Gogo-B and Patr-B alleles, and the unusual termination codon shared by Gogo-Okko and HLA-H. The alignment for the exon-5 region was obtained with the multiple sequence alignment program, TULLA (16). Exon boundaries are marked by short vertical bars. These sequence data have been submitted to the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers X60248-X60259, X60692, and X60693.



**Figure 2.** Pairwise comparison of gorilla (Gogo) class I alleles with HLA-A, -B, and -C alleles. (A) The frequency histogram obtained from all pairwise comparisons between 74 HLA-A, -B, and -C alleles. Differences from pairs involving alleles of different loci (A v B, A v C, B v C) are shaded solid; those from pairs of alleles from the same locus (A v A, B v B, C v C) are stippled. (B) The range of differences when the groups of Gogo-A, -B, or -C alleles are compared in a pairwise fashion to HLA-A, -B, or -C alleles.

it represents an allele of a locus that is distinct from Gogo-A and not the direct homologue of either HLA-A or HLA-H. Further insight into the unusual divergence of Gogo-Ok0 is gathered by an examination of the relationship of A-like sequences of the gorilla to individual families of HLA-A alleles.

On the basis of nucleotide sequence, all HLA-A alleles can be grouped into five families (A2, A3, A9, A10, and A19), and these correlate with the crossreacting groups of antigens defined by serology (21). Within families, pairs of HLA-A alleles differ by 2–32 nucleotide substitutions, compared to differences of 25–62 nucleotides between families (Fig. 5 a). These groupings also emerge from parsimony analysis, which additionally shows the five families to divide into two extended families: one comprising the A3 with A9 families; the other the A2, A10, and A19 families (Fig. 5 b).

To further define the families, we analyzed the HLA-A sequences for positions at which there are substitutions between families but homogeneity of nucleotides within the families. 33 such “family-specific” positions (two in exon 1, six in exon

HLA-A	GCAC	CCAA	TTT	CCAGTGGGCGTTA	CCTCTACCG	AAAAA	TGTCAT	TG
Gogo-A*0101	---	---	-c-	-----	g	-----	-----	---
Gogo-A*0201	---	---	---	-----	-----	-----	-----	---
Gogo-A*0401	---	---	---	-----	-----	-----	-----	---
Gogo-A*0501	-g--	---	-g	-----	a	-----	-----	---
Patr-A*01	---	---	-c-	-----	-----	-----	-----	---
Patr-A*02	---	---	---	-----	-----	-----	-----	---
Patr-A*03	---	---	-c-	-----	-----	-----	-----	---
Patr-A*04	---	---	---	-----	-----	-----	-----	---
Popy-A*01	-g--	---	-c-	-----	g	-----	-----	---
Gogo-Ok0	-g--	-t-g	g--	-----	g	-----	c--g	---
HLA-H	-gg-	-t-g	-c-	-----	g	-----	ca--g	---
HLA-E	c-	-t-g	gcc	gag--g--gg	-----	-t-	gg-gt	cc

HLA-B	CGGG	ATAA	GCC	CAGAAAAAATGA	TTCGAATAT	GGAGT	CGTCAA	CC
Gogo-B*0101	---	---	---	g--g--g--	-----	-----	-----	---
Gogo-B*0102	---	---	---	g--g--g--	-----	-----	-----	---
Gogo-B*0103	---	---	---	g--g--g--	-----	-----	-----	---
Gogo-B*0201	---	t--	---	g--gc-g	-----	tg-g	-----	---
Patr-B*01	---	---	---	g--g--g--	-----	tg-g	-----	---
Patr-B*03	---	---	---	ag--g--	-----	tg-g	-----	---
Patr-B*02	---	t--	t--	-----	-----	-----	-----	---
Patr-B*04	---	t--	t--	-----	-----	-----	-----	---
Patr-B*05	---	t--	t--	-----	-----	-----	-----	---

HLA-C	CGGG	CTGG	GCC	AAGGAAGAGGCGG	CCCGTTCT	GGGGT	CAATCT	CC
Gogo-C*0101	---	---	---	c-----	-t--aa-	-----	-----	---
Gogo-C*0102	---	---	---	c-----	-t--aa-	-----	-----	---
Gogo-C*0201	---	---	---	-----	-t-----	-----	-----	---
Gogo-C*0202	---	---	---	-----	-t-----	-----	-----	---
Gogo-C*0203	---	---	---	-----	-t-----	-----	-----	---
Patr-C*01	---	---	---	-----	-a-----	-----	-----	---

Exon	1	2	3	4	5	6	7	8
Position	4636	1122	344	6667777888888	999999991	11111	11111	11
	61	7922	468	2237788345667	134455570	00000	000000	11
	5868	585	1774737165174	037947920	12355	557889	00	
					2	91701	679895	01

**Figure 3.** Patterns of substitution for ape class I alleles at 46 locus-specific positions defined from comparisons of HLA-A, -B, and -C alleles. At each locus-specific position, all alleles of an individual HLA locus are identical, but differences exist between the loci. The number of these positions has been reduced from the 62 compared in a previous analysis (18), due to the inclusion of a recently discovered group of divergent HLA-C alleles. For each locus, the ape sequences are compared to the human consensus; dashes indicate identity. Sequences from gorilla (Gogo), chimpanzee (Patr), and orangutan (Popy) are included. The position numbers correspond to those of Fig. 1.

3, 14 in exon 4, and 11 in exons 5–8) were identified. Although only two alternative nucleotides are found at any of these positions, their combination provides an unequivocal fingerprint for each family (Fig. 6). The motifs at the family-specific positions further substantiate the similarities between the A3 and A9 families on the one hand and the A2, A9, and A10 families on the other. These observations suggest the five modern HLA-A families derive from two ancestral families: one that gave rise to the HLA-A3 and HLA-A9 families, the other to the HLA-A2, A10, and A19 families. Analysis of the pattern of substitution at the family-specific positions in the Gogo-A sequences substantiates this hypothesis.

Gogo-A\*0101, A\*0201, A\*0401, and A\*0501 have substitutions at family-specific positions characteristic of the HLA-A2, A10, A19 families and distinct from the HLA-A3 and A9 families. The gorilla motifs are a little more related to the HLA-A10 and -A19 family motifs than to the HLA-A2 family motif, but there is no clear segregation with either of the HLA-A10 or -A19 families. Thus, the Gogo-A\*0101, A\*0201, A\*0401, and A\*0501 sequences appear related to the common ancestor of the HLA-A10 and -A19 families, indicating these families evolved subsequent to the split of the human and gorilla lineages. The similarity of Gogo-A\*0101,

**Figure 4.** Comparison of part of the 3' untranslated region of gorilla (Gogo), chimpanzee (Patr), orangutan (Popy), and human (HLA) class I alleles. The nucleotides in bold letters are included to show the locus specificity present in this region.

# A

	1								90
Consensus	GACAGCTGCC	TTGTGTGGGA	CTGAGATGCA	GGAGTTTCTT	CACGCCCTCC	CCCCTTTGTG	ACTTCAAGAG	CCTCTGGCAT	CTCTTTCTGC
HLA-H	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----a-
Gogo-Ok0	-----	-----	-----g-	a-.-g-	-----t-	---.---t-	---g---a-	---a-t-	-----a-
Gogo-A*0101	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Gogo-A*0201	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Gogo-A*0401	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Gogo-A*0501	-----	-----	-----g-t	a-.-g-	-c---t-	t---.---t-	---g---a-	---aat-	-----a-
Popy-A*01	-----	-----	-----g-	a-.-g-	-c---t-	t---.---t-	---g---a-	---aat-	-----a-
Patr-A*02	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Patr-A*03	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Patr-A*04	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
A*0301	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
A*2601	-----	-----	-----g-	a-.-g-	-ct---t-	---c---t-	---g---a-	---a-t-	-----g-
A*0201	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
A*2401	-----	-----	-----g-	a-.-g-	-ct---t-	---.---t-	---g---a-	---a-t-	-----g-
Gogo-B*0101	-----tg	-----a-	-----	-----	-----	-----	-----	-----	-----
Gogo-B*0102	-----tg	-----a-	-----	-----	-----	-----	-----	-----	-----
Gogo-B*0103	-----tg	-----a-	-----	-----	-----	-----	-----	-----	-----
Gogo-B*0201	-----t-	-----a-	-----t-	-----	-----	-----	-----	-----	-----
Patr-B*01	-----t-	-----a-	-----	-----	-----	-----	-----	-----	-----
Patr-B*03	-----t-	-----a-	-----	-----	-----	-----	-----	-----	-----t-
Patr-B*02	-----t-	-----a-	-----a-	-----	-----	-----	-----	-----	-----
Patr-B*04	-----t-	-----a-	-----a-	-----	-----	-----	-----	-----	-----
Patr-B*05	-----t-	-----a-	-----a-	-----	-----	-----	-----	-----	-----
B*5801	-----t-	-----a-	-----	-----	-----	-----	-----	-----	-----
B*2701	-----t-	-----a-	-----	-----	-----	-----	-----	-----	-----
B*4401	-----t-	-----a-	-----	-----	-----	-----	-----	-----	-----
Gogo-C*0101	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Gogo-C*0102	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Gogo-C*0201	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Gogo-C*0202	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Gogo-C*0203	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Patr-C*01	-----	-----	-----	-----	-----a-	-----t-	-----c-a	-----	-----
Cw*0301	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Cw*0101	-----	-----	-----	-----c-	-----a-	-----t-	-----	-----	-----
Cw*0601	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----
Cw*0701	-----	-----	-----	-----	-----a-	-----t-	-----	-----	-----

# B

	91								179
Consensus	AAACGGCACC	TGAATGTGTC	TGCGTTCTCTG	TTAGGCATAA	TGTGAGGAGG	TGGAGAGAAC	CAGCCCACCC	CCAGTGTCCA	CCGTGACCC
HLA-H	-----	-----	-----t-	-----	-----t-	a-g-----	-----a-	t-a-----	-----a-
Gogo-Ok0	-----	-----	-----t-	-----c-	-----	ag-----	-----a-	---tcc-t-	-----a-
Gogo-A*0101	-----	-----c-	-----t-	-----	-----	g-----	-----c-	---a-----	-----a-
Gogo-A*0201	-----	-----c-	-----t-	-----	-----	g-----	-----c-	---a-----	-----a-
Gogo-A*0401	-----	-----c-	-----t-	-----	-----	g-----	-----c-	---a-----	-----a-
Gogo-A*0501	-----	-----c-	-----t-	-----c-g	-----	a-g-----	-----t-g-t-	-----	-----c-
Popy-A*01	-----	-----c-	-----t-	-----	-----	g-----	-----c-	---t-----	-----
Patr-A*02	-----	-----c-	-----t-	-----a-	-----t-	g-----	-----c-	---a-----	-----a-
Patr-A*03	-----	-----c-	-----t-	-----a-	-----t-	g-----	-----c-	---a-----	-----a-
Patr-A*04	-----	-----c-	-----t-	-----t-	-----	g-----	-----c-	---a-----	-----a-
A*0301	-----	-----c-	-----t-	-----g-	-----	g-----	-----c-	---a-----	-----a-
A*2601	-----	-----c-	-----t-	-----g-	-----	g-----	-----c-	---a-----	-----a-
A*0201	-----	-----c-	-----t-	-----g-	-----	g-----	-----c-	---a-----	-----a-
A*2401	-----	-----c-	-----t-	-----a-	-----	g-----	-----c-	---a-----	-----a-
Gogo-B*0101	-----	-----	-----a-c-	-----	-----	-----	-----	---t-----	-----
Gogo-B*0102	-----	-----	-----a-c-	-----	-----	-----	-----	---t-----	-----
Gogo-B*0103	-----	-----	-----a-c-	-----	-----	-----	-----	---t-----	-----
Gogo-B*0201	-----	-----	-----c-	-----	-----	-----	-----	---tt-----	-----t-
Patr-B*01	-----	-----	-----c-	-----	-----	-----	-----	---tt-----	-----t-
Patr-B*03	-----	-----	-----c-	-----	-----	-----	-----	---tt-----	-----t-
Patr-B*02	-----	-----	-----t-c-	-----	-----	-----	-----	---tt-----	-----t-
Patr-B*04	-----	-----	-----t-c-	-----	-----	-----	-----	---tt-----	-----t-
Patr-B*05	-----	-----	-----t-c-	-----	-----	-----	-----	---tt-----	-----t-
B*5801	-----	-----	-----c-	-----	-----	-----c-	-----	---tt-----	-----t-
B*2701	-----	-----	-----c-	-----	-----	-----	-----	---tt-----	-----t-
B*4401	-----	-----	-----c-	-----	-----	-----	-----	---tt-----	-----t-
Gogo-C*0101	-----t-	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-C*0102	-----t-	-----	-----	-----	-----	-----	-----	---a-----	-----
Gogo-C*0201	-----t-	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-C*0202	-----t-	-----	-----	-----	-----	-----	-----	-----	-----
Gogo-C*0203	-----t-	-----	-----c-	-----	-----	-----	-----	-----	-----
Patr-C*01	-----t-	-----	-----	-----	-----	-----	-----	---t-----	-----
Cw*0301	-----t-	-----	-----	-----	-----	-----	-----	-----	-----
Cw*0101	-----t-	-----	-----	-----	-----	-----	-----	-----	-----
Cw*0601	-----t-	-----c-	-----	-----	-----	-----	-----	-----	-----
Cw*0701	-----gt-	-----	-----	-----	-----	-----	-----	-----	-----

**Table 1.** Range of Nucleotide Differences with Gogo-Oko

	HLA-A	Gogo-A	HLA-H	Popy-A	Patr-A
Entire coding region	62-94*	69-85	76-81	78	68-75
Exon 1	5-8	3-6	9	6	6-7
Exon 2	24-40	26-40	12-20	29	30-34
Exon 3	12-25	15-23	29-34	27	17-19
Exon 4	7-12	8-11	10-11	6	9-11
Exon 5	4-9	7-8	7	6	4-5
Exon 6	1-3	1	1	1	0-2
Exons 7 and 8	2-3	2	3-5	3	0-2

\* The range of nucleotide differences when Gogo-Oko is compared to individual alleles of the designated class I loci. A single Popy-A allele was compared.

-A\*0201, -A\*0401, and -A\*0501 at both locus-specific and family-specific positions argues strongly that these cDNA sequences represent alleles of a gorilla Gogo-A locus that is the homologue of HLA-A.

In contrast, Gogo-Oko has a motif at the family-specific positions that is quite distinct from the Gogo-A sequences

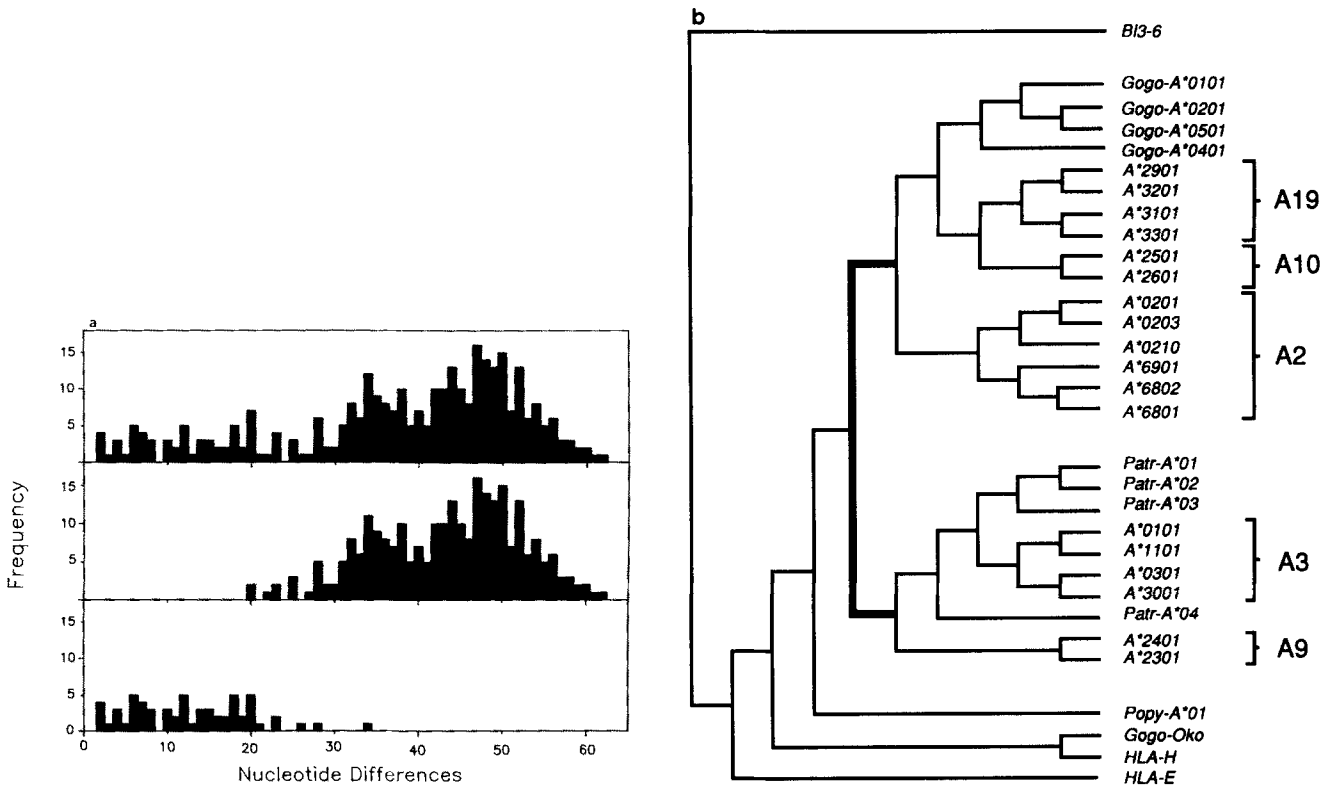
Consensus	CC	GCCCTG	GTCTGGCGGGG-C	CGGCAATTG	T	G
Patr-A*01	--	a-----	-c-ca-----ag-	-----c--	a	-
Patr-A*02	--	a-----	-c-ca-----ag-	-----c--	a	-
Patr-A*03	--	a-----	-c-ca-----ag-	-----c--	a	-
Patr-A*04	g-	a-----	-c-ca-----aag-	-----c--	a	-
A3 family	--	a-----	-c-ca-----g-	t-----c--	a	-
A9 family	g-	---a--	-c-ca-ta-aag-	a-----c--	-	-
A2 family	g-	---aca	a-g--a--a--ct	---t-----	-	-
A10 family	g-	-t-----	-g-----a-t-c-	---t-g--	-	a
A19 family	-t	-t-----	-g-----a-t-c-	---attg-g-	-	a
Gogo-A*0101	g-	-t-----	-g-----a--c-	-a-t-g-g-	-	-
Gogo-A*0201	-t	-t-----	-g-----a--c-	-a-t-g-g-	-	-
Gogo-A*0401	g-	-t-----	-g-----a--c-	-a-t-g-g-	-	-
Gogo-A*0501	--	-t-----	-g-----a--c-	-a-t-g-g-	-	-
Gogo-Oko	--	at-----	-----a--	-----	-	-
HLA-H	--	-----	-c-ca-----	-a-g-----	-	-
Popy-A*01	--	-----	-----a--a-	t-----	-	-
HLA-E	--	---t-a	---ca-----a-g-	---gt-----	-	.

Exon	1	3	4	5	6	7
Position	24	344445	66666666688888	899999991	1	1
	81	624590	34455677900126	904567990	0	0
		338376	02912628178998	905281381	3	7
					1	9
						8

**Figure 6.** Patterns of substitution for ape A alleles at family-specific positions defined on the basis of HLA-A alleles. The 33 positions define the five HLA-A allelic families; members of families possess identical nucleotides at these positions.

and that does not clearly segregate with any of the HLA-A families. In fact, its motif is closest to the consensus, suggesting this gene diverged before separation of any of the A families. This provides support for the possibility that Gogo-Oko is the product of a locus different from Gogo-A and



**Figure 5.** The relationship of ape A alleles to the different HLA-A families. Families of HLA-A alleles can be distinguished on the basis of overall sequence comparison. (a) In the upper panel are shown the frequency distributions of differences when all pairwise comparisons between HLA-A alleles are made. The middle panel shows the distribution for pairs of alleles derived from different families; the bottom panel shows the distribution for pairs of alleles from the same family. (b) A tree plotted from parsimony analysis (17) showing the segregation of HLA-A alleles into five families and two extended ancestral families. A bovine class I MHC sequence, B13-6, is the outlier for the analysis (36).



**Table 2. Shared Polymorphic Motifs**

Gorilla molecule	Similar molecule	Region of similarity
A*0101	A10s, A2s, A*2301, Patr-A*04	Leader peptide
	Popy-A*01	$\alpha 1$ (1–48)
	Gogo-A*0201	$\alpha 1$ (1–55)
	A*2301, A*2401	$\alpha 1$ (1–61)
	Patr-A*04	$\alpha 1$ (1–61)
	A*2501, A*2601	$\alpha 1$ (10–75)
	A*3301	$\alpha 2$ (10–72)
	Gogo-A*0501	$\alpha 1$ – $\alpha 2$ (84–170)
	Gogo-A*0201	$\alpha 2$ (98–162)
	A*3301, A*3101	$\alpha 2$ (98–140)
	B*1401, B*1402	$\alpha 2$ (153–181)
	Gogo-A*0401	$\alpha 3$ (2 distinctions)
	Gogo-A*0201	$\alpha 3$ (2 distinctions)
	A2s, A10s, A19s	$\alpha 3$ (few distinctions)
A*0201	A*2901, A*3201	Leader peptide
	Patr-A*04	$\alpha 1$ (1–55)
	A*2301, A*2401	$\alpha 1$ (1–55)
	A*0201, A*0203	$\alpha 1$ (10–73)
	Gogo-A*0101, -A*0501	$\alpha 2$ (98–162)
	A*3301, A*3101	$\alpha 2$ (98–140)
	A*0201	$\alpha 2$ – $\alpha 3$ (164–264)
A*0401	A10s, A2s	Leader peptide
	Patr-A*04	Leader peptide
	A*6801, A*6901	$\alpha 1$ (1–55)
	A*2601, A*2501	$\alpha 1$ (1–55)
	B*5701, B*5801	$\alpha 1$ (57–69)
	Gogo-A*0101	$\alpha 1$ – $\alpha 2$ (71–140)
	A*0201	$\alpha 3$ (3 distinctions)
A*0501	A10s, A2s, A*2301	Leader peptide
	Patr-A*04, A*2301	$\alpha 1$ (1–51)
	Patr-B*02, -04, -05	$\alpha 1$ (64–72)
	Bw4 +	$\alpha 1$ (79–83)
	A*2501	$\alpha 1$ – $\alpha 2$ (76–104)
	Gogo-A*0101	$\alpha 1$ – $\alpha 2$ (84–170)

**Table 2. (continued)**

Gorilla molecule	Similar molecule	Region of similarity
Gogo-Ok	Gogo-A*0201	$\alpha 2$ (98–162)
	A*3301	$\alpha 2$ (98–140)
	A*0201, A*6901	$\alpha 2$ – $\alpha 3$ (153–188)
	A*3201, A*2901	$\alpha 2$ – $\alpha 3$ (153–188)
	A2s, Gogo-A*0401	$\alpha 3$ (203–274)
	HLA-H	$\alpha 1$ (2–71)
	B*4201, B*0702	$\alpha 1$ (S24, E45, N63, Y67)
	A*2601	$\alpha 2$ (91–113)
	Gogo-A*0101	$\alpha 2$ (161–182)
	Popy-A*01	$\alpha 3$ (185–235)
	Gogo-C*0201, Cw*0701	tm (M307)
	HLA-H, HLA-E	tm (K311)
	B*0101, B*0102 & B*0103	Leader peptide
	Patr-B*03	$\alpha 1$ (1–64, 71–90)
B*0201	B*2702	$\alpha 1$ – $\alpha 2$ (71–94)
	Bw4 +	$\alpha 1$ (79–83)
	Patr-B*01	$\alpha 2$ (117–155)
	Patr-B*02, -B*03, -B*04	Leader peptide
	Patr-B*04	$\alpha 1$ (34–78)
	B*4101, B*4001	$\alpha 1$ (1–31)
	many HLA-Bs	$\alpha 1$ (50–75)
	A*2601	$\alpha 1$ – $\alpha 2$ (76–101)
	A*0101	$\alpha 1$ – $\alpha 2$ (76–96)
	Patr-B*03	$\alpha 2$ (117–155)
	B*5701	$\alpha 2$ – $\alpha 3$ (157–259)
	Patr-B*01, -B*03	tm and cyt. (275–339)
	Most HLA-Cs	Leader peptide
	Gogo-C*0201, Cw6p	$\alpha 1$ – $\alpha 2$ (44–94)
C*0101 & C*0102	B*5701	$\alpha 2$ – $\alpha 3$ (157–222)
	Cw*0701	cyt. (315–342)

**Table 2.** (continued)

Gorilla molecule	Similar molecule	Region of similarity
C*0201, C*0202 & C*0203	Most HLA-Cs	Leader peptide
	Cw*0601, Cw*0101	$\alpha 1$ (10–76)
	Cw6p	$\alpha 1$ - $\alpha 2$ (25–96)
	Cw*0701,	$\alpha 1$ - $\alpha 2$
	Cw*0702, HLA4	(81–113)
	Cw*0701	$\alpha 2$ - $\alpha 3$ (163–266)
	Cw*0701	tm and cyt. (295–342)

The gorilla molecules were assessed for regions of similarities with human, chimpanzee, and gorilla class I proteins. Numbers in parenthesis indicate the positions of the class I molecule where similarity is evident. tm, transmembrane; cyt., cytoplasmic.

that diverged from Gogo-A before diversification of the families.

That none of the Gogo-A sequences segregates with a single family of HLA-A alleles is a quite different situation from that observed for chimpanzee Patr-A sequences. In that species all the Patr-A alleles clearly show a close relationship with the HLA-A3 family. This can be seen both from overall comparisons and from the pattern of substitution at the family-specific positions. Interestingly, the presence of an A9-specific position at position 807 in Patr-A4 may belie the shared ancestry of the A3 and A9 families. Perhaps of greatest importance is the conclusion emerging from these analyses that the A locus molecules in gorillas and chimpanzees derived from two distinct ancestral lineages of A alleles, both of which are represented in humans.

**Proteins Motifs Shared by Gogo-A, Patr-A, and HLA-A Alleles.** As expected from the family relationships, none of the Gogo-A sequences is particularly similar to individual HLA-A alleles: the most similar pair being Gogo-A\*0101 and HLA-A\*6801, which differ at 27 nucleotide positions. Instead, they are formed of a pastiche of structural motifs, many of which can be recognized in class I HLA and Patr sequences (Table 2).

In pairwise comparisons, the Gogo-A\*0101, -A\*0201, and -A\*0401 proteins differs by 13–14 amino acid substitutions. Although distributed throughout the heavy chain, they are predominantly found in  $\alpha 1$  and  $\alpha 2$ , and in particular at peptide binding positions of those domains (Fig. 7). The number of differences and their distributions is comparable to that seen, for example, between HLA-A\*0201 and -A\*6801.

The  $\alpha 1$  domain of Gogo-A\*0201 is almost the same as that of HLA-A\*0201, and some similarities extend through the rest of these molecules. In particular, the motif at residues 62–65 of the  $\alpha 1$  helix, which is similar in HLA-A\*0201,

HLA-B\*5701, and -B\*5801 and gives rise to a shared alloantigenic site, is present (22). This motif is also present in Gogo-A\*0401, but interestingly in that case, the flanking amino acids are characteristic of those found in HLA-B\*5701 and -B\*5801.

The Gogo-A\*0101 protein has an  $\alpha 1$  domain and the first half of  $\alpha 2$ , which are clearly related to HLA-A\*3301. Gogo-A\*0501 differs from Gogo-A\*0101, -A\*0201, and -A\*0401 by 27–29 amino acids. Substitutions in the  $\alpha 1$  helix are the feature that make this allele distinct (17 of the 21 residues from positions 63–83 differ from those in Gogo-A\*0101, -A\*0201, and -A\*0401). The sequence in this region can be divided into three segments: positions 62–72, which are characteristic of motifs found in HLA-B but not HLA-A alleles; positions 73–76, which are unique; and positions 77–83, which comprise a motif common to HLA-A, -B, and -H alleles. The latter motif is responsible for the Bw4 public epitope found in certain HLA-A and -B molecules (23). In these B-like regions, Gogo-A\*0501 most closely resembles HLA-B\*4901.

**Gogo-B\*01: a Group of Closely Related Alleles.** The HLA-B locus is the most polymorphic human class I MHC locus. From analysis of four Gogo-B alleles from four unrelated gorillas, this same conclusion is not reached for its gorilla homologue. We obtained an identical allele, Gogo-B\*0103, from two of the individuals, and three of the four distinctive alleles (Gogo-B\*0101, B\*0102, B\*0103) only differ by one to four nucleotide substitutions. The pattern of substitutions between these three alleles is analogous to those found between closely related HLA-A or -B subtypes: all of the substitutions can be found in other class I Gogo alleles, all produce amino acid substitutions, and three of the four changes are at peptide binding residues of the antigen recognition site (Table 3). All these substitutions involve a localized cluster of positions on the first  $\beta$  strand of the  $\alpha 2$  domain; three (95, 97, and 99) have side chains that point up from the floor into the peptide binding groove and are among the sites with greatest variation in HLA-A, -B, and -C molecules (5, 24). The substitutions at these positions almost certainly result from natural selection for differences in antigen presentation. A striking example is provided by Gogo-B\*0101 and Gogo-

**Table 3.** Polymorphisms Distinguishing the Gogo-B\*01 Subtypes

	Position			
	*	*	*	
	95	97	99	103
Gogo-B*0101	I	R	F	V
Gogo-B*0102	F	R	F	V
Gogo-B*0103	I	W	Y	M
Gogo-B*0201	I	R	Y	M

Differences between the three subtypes of Gogo-B\*01. The residues at these positions for Gogo-B\*0201, a molecule that is divergent elsewhere, are also included as its motif is related to that found for Gogo-B\*0103.

\* Peptide-interacting residues.

**A****Leader**

	-24		-1
Consensus	MRVMAPRTLL	LLLSGALALT	ETWA
Popy-A*01	-a-----	-----	q---
HLA-H	-v-----	-----	q---
Gogo-Oko	-a-v-----	-t-----	r---
B*4201	-l-----v-	-a-----	---
A*2401	-a-g-----v	-----	q---
Patr-A*04	-a-----v	-----	q---
Patr-A*01	-a-p-----	-----	q---
Patr-A*02	-a-p-----	-----	q---
Patr-A*03	-a-p-----	-----	q---
A*2501	-a-----v	-----	q---
A*3301	ia-----	---l-----	q---
Gogo-A*0101	-a-----v	-----	q---
A*0201	-a-----v	-----	q---
B*1401	-l-----v-	-a-----	---
Gogo-A*0201	-a-----	---l-----	q---
Gogo-A*0401	-a-----v	-----	q---
B*5701	---t-----	---w-v---	---
Gogo-A*0501	-a-v-----	-----	q---
B*4901	---t-----v-	-a-----	---
Patr-B*02	-q-t-----v-	-a-----	---
Patr-B*03	-q-t-----v-	-a-----	---
Gogo-B*0101	---t-----	-a-----	---
Gogo-B*0102	---t-----	-a-----	---
Gogo-B*0103	---t-----	-a-----	---
B*2702	---t-----	---w-v---	---
B*4001	---t-----v-	-a-----	---
Gogo-B*0201	-q-t-----	-a-----	---
A*0101	-a-----	-----	q---
Gogo-C*0101	-----i	-----	---
Gogo-C*0102	-----i	-----	---
Gogo-C*0201	-----i	-p-----	---
Gogo-C*0202	-----i	-----	---
Gogo-C*0203	-----i	-----	---
Cw6p	.....	.....	---
Cw*0601	-----i	-----	---
Consensus	MRVMAPRTLL	LLLSGALALT	ETWA

 **$\alpha$ 1 domain**

Accession	1		31		61		90		
Consensus	GSMSMRIFYT	SVSRPGRGEP	RFIAVGIVDD	TQFVRFDSDA	ASPRMEPRAP	WIEQEGPEYW	DRETQIVKAQ	SQTDRLVSLRT	LRGYYNQSEA
Popy-A*01	-----s	-----	-----	-----	-q-----t	-m-----	---rs---h	a-n-d-g-	-----dg
HLA-H	r-----	tm-----	---s-----	-----d	-e-----	-m-r-----	---n-c---	ar-e-en-i	alr-----g
Gogo-Oko	-----	tm-----	---s-----	-----d	-e-----	-m-r-----	---n-y---	a---d- <del>e</del>	-----g
B*4201	-----s	-----	---s-----	-----	-e-----	-----	---n-y---	a---e-en	-----
A*2401	-----s	-----	-----	-----	-q-----	-----	-e-gk---h	---en-i	alr-----
Patr-A*04	-----s	-----	-----	-----	-q-----	-----	-e-rs---s	a---d-g-	-----d
Patr-A*01	-----f	-----	-----	-----	-q-----	-----	-q-rsa-h	---d-g-	-----d
Patr-A*02	-----f	-----	-----	-----	-q-----	-----	-e-rsa-h	---d-g-	-----d
Patr-A*03	-----	-----	-----	-----	-q-----	-----	-q-rnm-s	a---d-g-	-----d
A*2501	-----	-----	-----	-----	-q-----	-----	---n-rn---h	---e-i	alr-----d
A*3301	-----t	-----	-----	-----	-q-----	-----	---n-rn---h	-i-d-g-	-----d
Gogo-A*0101	-----s	-----	---n-rn---	-----d	-q-----	-----	---n-rn---h	---d-g-	-----d
A*0201	-----f	-----	-----	-----	-q-----	-----	-g-rk---h	---h-d-g-	-----d
B*1401	-----	-----	---s-----	-----	-e-----	-----	---n-c-tn	t---e-n	-----
Gogo-A*0201	-----s	-----	-----	-----	-q-----	-----	-g-rk---h	---n-g-	-----
Gogo-A*0401	-----	-----	-----	-----	-q-----	-----	-g-rnm-r	---d-g-	-----d
B*5701	-----	am-----	-----	-----	-q-----a	-----	-g-rnm-s	a-y-en-i	alr-----d
Gogo-A*0501	-----s	-----	-----	-----	-q-----	-m-----	-q---s-tn	a-iele-i	alr-----d
B*4901	-----h	am-----	---t-----	-l-----	t---k-----	---l-----	---s-tn	t-y-en-i	alr-----
Patr-B*02	-----k	a-----	---s-----	---w-----	t---e-----	-----	---s-tn	a-y-en-n	-----
Patr-B*03	-----d	a-----	---t-----	-----	-----	-----	---rn---s	a---en-i	alr-----
Gogo-B*0101	-----d	a-----	---t-----	-----	-----	-----	---ts---	a---en-i	alr-----
Gogo-B*0102	-----d	a-----	---t-----	-----	-----	-----	---ts---	a---en-i	alr-----
Gogo-B*0103	-----d	a-----	---t-----	-----	-----	-----	---ts---	a---en-i	alr-----
B*2702	-----h	-----	---t-----	-l-----	---e-----	-----	---c-k---	a---en-i	alr-----
Patr-B*04	.....	.....	.....	---w-----	---k-t---	-----	---s-tn	a-y-g-n	---lr-----
B*4001	-----h	am-----	---t-----	-l-----	t---k-----	-----	---s-tn	t-y-e-n	-----
Gogo-B*0201	-----h	am-----	---t-----	---w-----	---k-t---	-----	---s-tn	t-y-g-g-	-----d
A*0101	-----f	-----	-----	-----	-qk-----	-----	-q-rnm-h	---an-g-	-----d
Gogo-C*0101	-----f	a-ap-----	-----	-----	-nt-g-----	-v-----	---ky-r-	a-----k	-----d
Gogo-C*0102	-----f	a-ap-----	-----	-----	-nt-g-----	-v-----	---ky-r-	a-----k	-----d
Gogo-C*0201	-----	-----	---s-----	-----	-g-----	-v-----	---ky-r-	a-----n-k	-----d
Gogo-C*0202	-----	-----	---s-----	-----	-g-----	-v-----	---ky-r-	a-----n-k	-----d
Gogo-C*0203	-----	-----	---s-----	-----	-g-----	-v-----	---ky-r-	a-----n-k	-----d
Cw6p	c-----c	a-----	---s-----	-----	-g-----	-v-----	---ky-r-	a-a-n-k	-----d
Cw*0601	c-----	a-----	---s-----	-----	-g-----	-v-----	---ky-r-	a-a-n-n	-----d
Cw*0701	.....	.....	.....	.....	.....	.....	---ny-r-	a-a-n-n	-----d
Consensus	GSMSMRIFYT	SVSRPGRGEP	RFIAVGIVDD	TQFVRFDSDA	ASPRMEPRAP	WIEQEGPEYW	DRETQIVKAQ	SQTDRLVSLRT	LRGYYNQSEA

# B

## α2 domain

	91			121			151		182	
Consensus	GSHTLQRMYG	CDVGPDPGRLL	RGYDQDAYDG	KDYIALNEDL	RSWTAADTAA	QITQRKWEAA	RVAEQLRAYL	EGTCVEWLRR	YLENGKETLQ	RA
Popy-A*01	---i---f	---f---f	---e-h---	---	---m---	---	ga---d---	---l---s---	---	-t
HLA-H	---m-v---	---l---f---	---e-h---	---	---m---	---k---	---r---r-v---	---ef---	---	---
Gogo-Oko	---i---	---e---f---	---l---	---t---	---	---	---e---r---m---	---	h---	-t
B*4201	---s---	---	---hn-y---	---	---	---	---d---	---	---d-e---	---
A*2401	---m-f---	---s-f---	---h-y---	---k---	---m---	---k---	h---q---	---dg---	---	---
Patr-A*04	---i-l-f---	---s-f---	---r---	---	---m---	---	ha---	---	---	-t
Patr-A*01	---i-i---	---s-f---	---r---	---	---m---	---k---	ha---q---	---	---	-t
Patr-A*02	---i-i---	---s-f---	---r---	---	---m---	---k---	ha---r---	---	---	-t
Patr-A*03	---i-i---	---s-f---	---r---	---	am---	---k---	ha---	---r---	---	-t
A*2501	---i---	---f---	---q---	---	---	---	he---w---	---r---	---	-t
A*3301	---i-m---	---s-f---	---q---	---	---	---	---de---	---	h---	-t
Gogo-A*0101	---i---	---s-f---	---q---	---	---	●-k---	hf---	---	h---	-t
A*0201	---v---	---s-w-f---	---h-y---	---k---	---	h- -t-kh-	h-	---	---	-t
B*1401	---w---	---	---n-f---	---	s---	---	---e---	---	h---	---
Gogo-A*0201	---i-k---	---s-f---	---q---	---	---	●-k---	hf---	---e---	---	-t
Gogo-A*0401	---i---	---s-f---	---q---	---	---	---	he---	---	---	lt
B*5701	---i-v---	---	---h-s---	---	s---	---	---	---l---	---	---
Gogo-A*0501	---i---	---s-f---	---q---	---	---	●-k---	hf---	---	---	-t
B*4901	---w---	---l---	---n-l---	---	s---	---	---e---	---l---	---	---
Patr-B*02	---i-i---	---m---	---e-y---	---	s---	---	---w---	---	---	-v
Patr-B*03	---w-t-f---	---m---	---y-y---	---	s---	---	---a---	---	---	---
Gogo-B*0101	---i---	---f---	---s---	---	s---	---	---e---	---	---	-r
Gogo-B*0102	---f---	---f---	---s---	---	s---	---	---e---	---	---	-r
Gogo-B*0103	---i-w---	---m---	---s---	---	s---	---	---e---	---	---	-r
B*2702	---n---	---	---h---	---	s---	---	---	---e---	---	---
Patr-B*04	---w-t-f---	---	---r-y---	---q---	s---	---	---	---	---	---
B*4001	---	---	---hn-y---	---	---	s-l	---	---e---	---	-dk-e
Gogo-B*0201	---i---	---m---	---s---	---l---	s---	---	a-●	---l---	---	---
A*0101	---i-i---	---f---	---r---	---	---	---	ha---r-v---	---r-dg---	---	-t
Gogo-C*0101	---f---	---q---	---l---	---	---	---	---w-rq---	---l---	---	---
Gogo-C*0102	---f---	---q---	---	---	---	---	---	---	---	---
Gogo-C*0201	---s---	---l---	---s-f---	---	---	v---	---e---	---l---	---	---
Gogo-C*0202	---s---	---l---	---s-f---	---	---	l---	a-q	---l-s---	---	---
Gogo-C*0203	---s---	---l---	---s-f---	---	---	v---	---e---	---l---	---	---
Cw6p	---w---	---l---	---s---	---	---	---	---e-w---	---	---	---
Cw*0601	---w---	---l---	---s---	---	---	---	a-q	---	---	---
Cw*0701	---	---l---	---s---	---	---	l	---	---	---	---
Consensus	GSHTLQRMYG	CDVGPDPGRLL	RGYDQDAYDG	KDYIALNEDL	RSWTAADTAA	QITQRKWEAA	RVAEQLRAYL	EGTCVEWLRR	YLENGKETLQ	RA

## α3 domain

	183			213			243		274	
Consensus	DPPKTHVTHH	PVSDHEATLR	CWALGFYPAE	ITLTWQRDGE	DQTQDTLVE	TRPAGDGTFO	KWAAVVVPSPG	EEQRYTCHVQ	HEGLPKPLTL	RW
Popy-A*01	-a--m-							k-	-e-	
HLA-H	-a--m-	-i-			-hx-			k-	-e-	
Gogo-Ok0	-m-					-g-		k-		
B*4201	-m-	-i-				-r-				
A*2401	-m-	-i-								
Patr-A*04	-m-	-i-r								
Patr-A*01	-m-	-i-								
Patr-A*02	-m-	-i-								
Patr-A*03	-m-	-i-	-v							
A*2501	-a--m-	a-	-s-				-s-	q-		
A*3301	-r-m-	a-	-s-				-s-	q-		
Gogo-A*0101	-a--m-	a-	-i-	-s-				q-	-e-	
A*0201	-a--m-	a-	-s-					q-		
B*1401		-i-				-r-				
Gogo-A*0201	-a--m-	a-	-i-	-s-				q-	-a-	
Gogo-A*0401	-a--m-	-i-	-s-					q-		
B*5701		-i-				-r-				
Gogo-A*0501	-a--t-q	a-	-k-	-s-				q-		
B*4901		-i-				-r-				
Patr-B*02		-i-				-r-				
Patr-B*03	-m-	-i-				-e-r-				
Gogo-B*0101	-t-	-i-						-e-		
Gogo-B*0102	-t-	-i-						-e-		
Gogo-B*0103	-t-	-i-						-e-		
B*2702		-i-				-r-				
Patr-B*04		-i-				-r-				
B*4001		-i-				-r-				
Gogo-B*0201		-i-				-r-		-r-	-e-	
A*0101	-m-	-i-				-r-				
Gogo-C*0101		-i-			e-i-		-m-		-k-le	
Gogo-C*0102		-i-			e-i-		-m-		-k-le	
Gogo-C*0201	e-	-l-						q-m-	-e-	
Gogo-C*0202	e-	-l-						q-m-	-e-	
Gogo-C*0203	e-	-l-						q-m-	-e-	
Cw6p	eh-								-e-	
Cw*0601	eh-	-l-							-e-	
Cw*0701	e-	-l-						q-m-	-e-s-	
Consensus	DPPKTHVTHH	PVSDHEATLR	CWALGFYPAE	ITLTWQRDGE	DQTQDTLVE	TRPAGDGTFO	KWAAVVVPSPG	EEQRYTCHVQ	HEGLPKPLTL	RW

## C

Transmembrane				Cytoplasmic domains			
	275		305				342
Consensus	EPSSQPTIPI	VGIVAGLA VL	AVVVI.GAVV	AAVMCRKSS	GGKGGYSQA	ASSDSAQGS D	VSLTACKV
Popy-A*01	-l-----	---i---vl-	ga-it----	---w-rn-	dr-----	--n-----	-----
HLA-H	-----	---i---vl-	va-t-----	---w-k---	dr-----	--gn-----	-----
Gogo-Oko	-----	---i---vl-	ga-it----	---m-w-k---	-r-----	-----	-----
B*4201	-----	---s-v---	-----	-----	-----	-c-----	-----
A*2401	-----	---v---	---i---vl-	ga-it----	---w-n-	dr-----	-----
Patr-A*04	-----	-----	---i---vl-	ga-it----	---w---	dr-----	-----
Patr-A*01	-----	-----	---i---vl-	ga-it----	---w....	-----	-----
Patr-A*02	-----	-----	---i---vl-	ga-it----	---w---	dr-----	t-----
Patr-A*03	-----	-----	---i---vl-	ga-it----	---w---	dr-----	t-----
A*2501	-----	-----	---i---vlf	ga-ia-----	---w---	dr-----	m-----
A*3301	-----	-----	---i---vlf	ga-fa-----	---rw---	dr-----	f-----
Gogo-A*0101	-----	-----	---i---vlf	ga-ia-----	---rw---	dr-----	m-----
A*0201	-----	-----	---i---vlf	ga-it----	---w---	dr-----	-----
B*1401	-----	---s-v---	-----	-----	-----	-----	-----
Gogo-A*0201	-----	-----	---i---vlf	ga-ia-----i	---rw---	dr-----	-----
Gogo-A*0401	-----	-----	---i---vlf	ga-ia-----	---rw---	dr-----	-----
B*5701	-----	---s-v---	-----	-----	-----	-c-----	-----
Gogo-A*0501	-----	-----	---i---vlf	ga-ia-----	---rw---	dr-----	-----
B*4901	-----	---s---	-----	-t-----	-----	-----	-----
Patr-B*02	-----	---s---	-----	-----	-----	-----	-----
Patr-B*03	-----	---s---	v-t-avv---	---l---	-----	-----	-----
Gogo-B*0101	-----	---s---	-----	t-i-----	-----	-----	-----
Gogo-B*0102	-----	---s---	-----	t-i-----	-----	-----	-----
Gogo-B*0103	-----	---s---	-----	t-i-----	-----	-----	-----
B*2702	-----	---s-v---	-----	-----	-----	-c-----	-----
Patr-B*04	-----	---s---	-----	-----	-----	-----	-----
B*4001	-----	---s-v---	-----	-----	-----	-c-----	-----
Gogo-B*0201	-----	---s---	v-t-avv---	-----	-----	-----	-----
A*0101	-l-----	---i---vl-	ga-it----	---w---	dr-----	-t-----	-----
Gogo-C*0101	-----	-----	---ft-t---	-----	-----	-c-n-----	e-i--a
Gogo-C*0102	-----	-----	---ft-t---	-----	-----	-c-n-----	e-i--a
Gogo-C*0201	-----	---v---	v-lavl---	t-m-----	-----	-c-n-----	e-it-a
Gogo-C*0202	-----	---v---	v-lavl---	t-m-----	-----	-c-n-----	e-it-a
Gogo-C*0203	-----	---v---	v-lavl---	t-m-----	-----	-c-n-----	e-it-a
Cw6p	-----	-----	---lavl-m	-v-----	-----	-c-n-----	e-i--a
Cw*0601	-----	-----	---lavl---	-v-----	-----	-c-n-----	e-i--a
Cw*0701	-----	-----	v-lavl---	t-m-----	-----	-c-n-----	e-it-a
Consensus	EPSSQPTIPI	VGIVAGLA VL	AVVVI.GAVV	AAVMCRKSS	GGKGGYSQA	ASSDSAQGS D	VSLTACKV

**Figure 7.** Comparison of the predicted protein sequences of gorilla class I heavy chains with class I heavy chains from other apes and humans. Dashes indicate identity with the HLA consensus. Amino acid substitutions found in gorilla molecules but not in human class I proteins are in bold.

B\*0102, which differ by a single nucleotide that substitutes phenylalanine for isoleucine at position 95.

The one substitution that is not at a peptide binding position, methionine at position 103 in Gogo-B\*0103, is associated with substitutions at two neighboring peptide-binding residues (97 and 99), and its presence may be the result of a “hitchhiking” effect in which the product of a recombinational event was selected because of the substitutions at positions 97 and 99. Alternatively, this substitution may have functional effects not yet appreciated.

Their close relationship strongly indicates that Gogo-B\*0101, -B\*0102, and -B\*0103 derive from a common ancestor. The most related class I gene is HLA-B\*4901, which differs at 40 or 41 nucleotide positions. Comparison with other class I alleles reveals elements of sequence that are shared with human and chimpanzee molecules (Table 2). For most of the leader peptide and residues 1–64 of the  $\alpha_1$  domain, Gogo-B\*0101, -B\*0102, and -B\*0103 are identical to the chimpanzee B locus molecule, Patr-B\*03. The COOH-terminal part of the  $\alpha_1$  domain, residues 70–90, is identical to HLA-B\*2702, a feature that may explain the high frequency with which HLA-B27 alloantisera react with gorillas. Residues 65–70 of the  $\alpha_1$  helix have a unique motif not seen in other class I molecules.

Characteristically, it is difficult to identify evolutionary relationships in the  $\alpha_2$  domain of HLA-B due to the extensive reassortment of substitutions in different molecules. The same is true for Gogo-B. In the COOH-terminal, two-thirds of  $\alpha_2$ , Gogo-B\*0101, -B\*0102, and -B\*0103 are similar to HLA-B\*1401, the differences being due to one substitution

that is unique to Gogo-B\*0101, -B\*0102, and -B\*0103 (arginine at 176) and another (histidine 171) shared by HLA-B\*1401 and HLA-B5 molecules (Table 2).

**Gogo-B\*0201: a Divergent Allele.** The Gogo-B\*0201 allele differs from the Gogo-B\*01 group by 58–60 nucleotide substitutions that give rise to 36–38 amino acid differences. These include 15–16 differences in peptide binding or TCR-interacting residues of the antigen recognition site. The Gogo-B\*0201 heavy chain bears elements in common with Patr-B\*03 and the HLA-B40 group of alleles in the NH<sub>2</sub>-terminal 76 residues of the  $\alpha_1$  domain. The COOH-terminal part of  $\alpha_1$  has a motif not found in the HLA-B alleles. This region produces the alternative Bw4 or Bw6 epitopes, one of which is found on all HLA-B molecules. The Gogo-B\*0201 molecule has neither motif in this region, instead having a sequence that is characteristic of HLA-A molecules.

HLA-B alleles are characterized by their homogeneity of sequence in exons 4–8, a feature that is not so strong for the Gogo-B alleles (18). Gogo-B\*0201 differs by five amino acids from Gogo-B\*01 in  $\alpha_3$  and has a distinctive motif with an additional amino acid in the transmembrane region. This transmembrane region polymorphism is also seen in two Patr-B alleles, Patr-B\*01 and -B\*03 (10, 11), and its origins may thus predate the divergence of the ape species (Fig. 1).

**Preservation of Two Divergent Groups of C Alleles in Gorillas and Humans.** The pattern of highly related alleles seen at the Gogo-B locus is reiterated with the five Gogo-C alleles. They divide into two groups: one consisting of Gogo-C\*0101 and -C\*0102; the other of Gogo-C\*0201, -C\*0202, and



including the two disulphide bonds and the single glycosylation site at position 86. There is every indication from comparison with the HLA-A, -B, and -C protein sequences that all the Gogo class I heavy chains will assemble with  $\beta_2$ -microglobulin ( $\beta_2$ -m) and peptide to form molecules that present antigens to T cells. Previously, we found gorilla  $\beta_2$ -m to be identical in protein sequence to both human and chimpanzee  $\beta_2$ -m (12).

Our previous comparison of chimpanzee and human class I genes showed that a majority of the substitutions found at positions of variation in the chimpanzee alleles were also found in human alleles, and that extended motifs of sequences were preserved in the two species (12, 18, 29). This pattern is consistent with a scenario in which the lifetime of alleles and polymorphic motifs can be longer than those of species and in which new species can inherit multiple alleles from ancestral species. A survey of the variation in the Gogo-A, -B, and -C alleles give the same result: most of the polymorphic substitutions can be found in HLA-A, -B, and -C alleles (78%), Patr-A, -B, and -C alleles (51%) or either (82%). There are, however, substitutions that are unique to the gorilla lineage (Fig. 9), and as these frequently lead to amino acid replacements in the antigen recognition site, they are likely to be the result of natural selection for novel antigen-presenting properties. The number of these unique substitutions found in the gorilla class I heavy chains significantly exceeds that seen for the chimpanzee heavy chains (Fig. 9).

## Discussion

Gogo-A*0101	-----	-----	-----E-F-----	I-----	-----
Gogo-A*0201	-----	-----E-----	-K-----E-F-----	I-----S-----	I-----
Gogo-A*0401	-----	-----E-R-----	-----L-----	I-----	-----
Gogo-A*0501	-V-----	-----ME-----EL-----	-----S-----	-TQ-K-----	-----
Gogo-Oko	-V--TR T-----	M-----EG-----	-----L-T-----R-M-----	-----G-K-----	-----K-----
Gogo-B*0101	-----	-----T-----	-----S-----R-----	T-----E-----	-----
Gogo-B*0102	-----	-----T-----	F-----S-----R-----	T-----E-----	-----
Gogo-B*0103	-----	-----T-----	-----M-S-----R-----	T-----E-----	-----
Gogo-B*0201	Q-----	-----W-T-----G-----	-----M-S-L-----E-----	-----R-----	TVV-----
Gogo-C*0101	-----	-----AP-NT-----	F-----Q-----R-----	-----I-M-K-L-----	T-----
Gogo-C*0102	-----	-----AP-NT-----	F-----Q-----R-----	-----M-K-L-----	T-----
Gogo-C*0201	-----P-----	-----	-----S-----V-E-----	-----	V-----
Gogo-C*0202	-----	-----	-----S-----	-----	V-----
Gogo-C*0203	-----	-----	-----S-----V-E-----	-----	V-----
Patr-A*01	--P----	-----SA-----	-----	-----	-----
Patr-A*02	--P----	-----SA-----	-----	-----	-----
Patr-A*03	--P----	-----	-----A-----	-----V-----	-----
Patr-A*04	-----	-----S-----	-L-----	-----R-----	-----
Patr-B*01	-----	-----	-----M-G-----V-----	-----E-----	TVV--L-----
Patr-B*02	Q-----	-----W-----	-----M-----V-----	-----	-----
Patr-B*03	Q-----	-----	-----M-Y-----	-----E-----	TVV--L-----
Patr-B*04	-----	-----W-T-----G-----	-----Q-----	-----	-----
Patr-B*05	Q-----	-----W-----	-----M-----Q-----	-----	-----
Patr-C*01	-----	-----H-----	-----Y-----	-----	A-----
Position	222194	111344455666777779	99111111111111111111	111122222222222222	2223333333
	3103	134523926167045790	570111222234455566788	88990001234556666	889000001
			3049456871225606612	49270242867350457	897012481
	leader	$\alpha 1$	$\alpha 2$	$\alpha 3$	tm

	I	II	III
	1 29	1 18	1 7
Consensus	MYFFAGMYEISQTDNTLYIRYDDTWELYY	PDQKAQVRRRAAREAGER	ERKQTW
Gogo-A*0101	--s-----nnvh--d-----q--f--h	--r---g---h-----	--r---g---h-----
Gogo-A*0201	--s-----kvh-----k-q--f--	--r---g---h-----	--g---e-----
Gogo-A*0401	-----nmr--d-----q-----	--r---g---h-----	--g---e-----
Gogo-A*0501	--s-----q--niesia--q--f--	--t-e-l--h-----	-----
Gogo-Ok0	-----s-e-n-y--d-----l--h	-----e-----	--r-----
A*2901	--t-----qnv-----m-r--v--	--r---ag-----	--l-----
A*3301	--t-----nnvhi--d-----m-q--v--h	--r---g---d-----	--e-----
A*2501	-----nnvh--sia--q--w--	--r---e-l-t-h-----	--r-----
A*0201	--f-----kvh--hd--v--hy--v--	--r---g-h--h-----	--g-----
A*0101	--f-----nmh-----i-r--ar--	--r---ag--vh-v-d--	--q--rg--
A*3001	--f-----nv--d-----i-eh--w--	--r---g-----	--q-----
A*2401	--s-----kvh--ia--lmfhy--vq--	--g---e-l--h--d--	--e--g--
Patr-A*01	--f-----sah--d-----i-r--aq--	--r---g---h-----	--q-----
Patr-A*02	--f-----sah--d-----i-r--ar--	--r---g---h-----	--e-----
Patr-A*03	-----nms--d-----i-r--a--	--r---g---h-----	--q--r--
Patr-A*04	--s-----svs--d-----lfr--a--	--r---g---h-----	--e-----
Popy-A*01	--s-----svh--nd-----feh--ad--	--r---g---g-----	--ls--
Gogo-B*0101	--d-t---t-----ia--fss-----	-----e-l-----	-----
Gogo-B*0102	--d-t---t-----ia-f-fss-----	-----e-l-----	-----
Gogo-B*0103	--d-t---t-----ia--w--ss-----	-----e-l-----	-----
Gogo-B*0201	--h-t-k---n-yg-----sl--ae--	-----t-g-----	--l-----
B*0702	-----s-e-n-y--sn--ls--y--r--	-----e-----	--e-----
B*1401	-----s-e-n-cn--sn--lw-nf-----h	--t-e-----	-----
B*2702	--h-t-e---ck---ia--ln-h--v--	-----e-l-----	--e-----
B*4001	--h-t-k---n-ysn--l--nyslv--	-----t-e-----	--e-----
B*4402	--t-k---n-y--a-----vd--	--t-e-l-----	--ls--
B*4701	--t-k---n-yd--l-fh--v--	--t-e-l-----	--e-----
B*4901	--h-t-k---n-y-ia-w--nl-----	--t-e-l-----	--l-----
B*5701	-----nms-y-ia--v--s--v--	--r---e-l-----	--g--l--
Patr-B*01	--t-----nms--ia--wt-gy--r--	--r---e-l-----	-----
Patr-B*02	-----s-e---n-ysn-----ey--w--	--t-e-----	-----
Patr-B*03	--d-t---nvs--ia--wtffy--a--	--r---e-l-----	-----
Patr-B*04	.....k---n-ygn--wtffy--v--	--t-e-l-----	-----
Patr-B*05	-----s-e---n-ysn-----ey--w--	--t-e-----	-----
Gogo-C*0101	--f---g--ky---sk--f--l--wq--	--r-----	--rl--
Gogo-C*0201	--s-g--ky---k--ls-sf--ve--	--r-----	--l-----
Gogo-C*0202	--s-g--ky---k--ls-sf-la-q--	--r-----	--ls--
Gogo-C*0203	--s-g--ky---k--ls-sf--ve--	--r-----	--l-----
Cw*0101	--f-s-g--ky---sn--lwc-y--r--	--r-----	-----
Cw*0201	-----g--ky---k--l--s--w--	--r-----	--e-----
BewoC-1	--s---g--ky-a--k--l-fnf--r--	--nr-----	--e-----
Jy328	--d-s-g--ky-a--sn--l-s-s-la--	-----d-----	--e-----
Patr-C*01	.....g--ky-a--sn--lw--y-la--	--h-r-----	-----
Consensus	MYFFAGMYEISQTDNTLYIRYDDTWELYY	PDQKAQVRRRAAREAGER	ERKQTW
Position	57922245666777788899911111111 24659367034701457911445557 46372691	5666677781111111111 715892692445555666 590148269	5611111 8245667 63370

**Figure 10.** Unique patterns of amino acid substitution at the antigen recognition sites of gorilla, chimpanzee, and human class I molecules. Only residues at the functional positions of the antigen recognition site as defined by the crystallographic structure are shown (24, 37). They are separated into putative peptide binding residues (I), TCR-contacting residues (II), and residues that may be involved in either or both of these contacts (III). Numbering corresponds to that in Fig. 7.

related has been highly controversial and remains unresolved (30, 31). At present, the most common view is that chimpanzees and humans share a more recent common ancestor than either species does with the gorilla (7–9). It is to be expected that final resolution of the trichotomy will come with comparison of extensive segments of homologous DNA that are subjected to minimal or no selection (32).

DNA encoding classical class I heavy chains is not of this type. These sequences are selected for antigen-presenting function and much of their variation and polymorphism can be attributed to that selection (5, 33). Within species there is variation between class I alleles that can be far greater than seen between species for many other genes. Furthermore, it is clear that the number of class I genes and their patterns of expression change on the evolutionary timescale (4). Alleles at a particular class I locus may number into the hundreds and their distribution within a population can be quite even (34). Factors that may have contributed to differences in the

class I alleles and loci of gorillas, chimpanzees, and humans are the pools of alleles and haplotypes inherited by new species from ancestral species, the nature and course of selection along different lineages, and neutral divergence.

In humans, HLA-A, -B, and -C are the class I loci implicated in antigen presentation and T cell immunity (35). Homologues of each of the three loci are expressed in the gorilla and chimpanzee, and the similarities between the homologous loci in the three species are such that no species-specific patterns of substitution can be discerned. For the chimpanzee class I sequences so far analyzed, all have conformed with the patterns established from analysis of class I HLA genes. This is not the case for the gorilla class I alleles, which in various ways are seen to break out of, or extend, the HLA pattern. For example, the number of unique substitutions found with the gorilla alleles significantly exceeds those found with the chimpanzee, and there is some unusual mixing of polymorphic motifs between the loci.

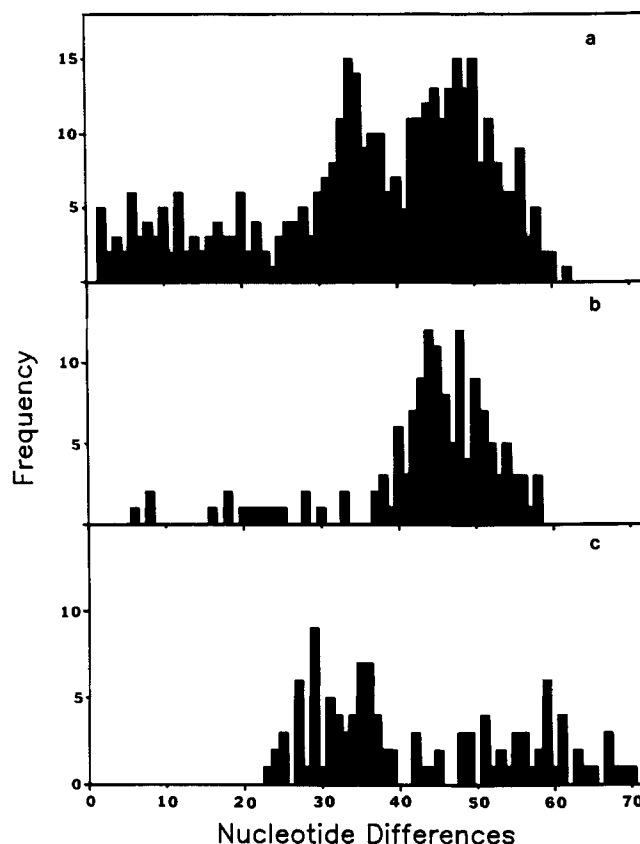


Most striking is the Gogo-Okoko gene. Although more related to HLA-A than to either HLA-B or -C, it is not clear if this gene is an allele of Gogo-A, the gorilla homologue of the HLA-A locus. It has extended sequence identities with HLA-H, an HLA-A-related pseudogene, and does not segregate with any of the established families of A alleles. Watkins et al. (20) have isolated a similar clone to Gogo-Okoko and suggest it was formed by a segmental exchange that introduced a segment of an HLA-H-like gene into an A allele. Our analysis does not favor this interpretation over plausible alternatives, in which Gogo-Okoko is an allele of a locus distinct from Gogo-A. One possibility is for this locus to be the homologue of HLA-H, in which case the divergence of their sequences might be attributed to the profound difference that in the human lineage this gene (HLA-H) became nonfunctional, whereas in the gorilla lineage it may have retained function. A second possibility is that Gogo-Okoko represents a third A-related gene that has not been described, or does not exist, in humans.

A second feature that distinguishes the gorilla class I loci is in differences between the A and B loci. By various parameters (the numbers of alleles, their relative frequency, and the diversity of the antigen recognition site), HLA-B appears a more functionally diversified locus than HLA-A (5). Furthermore, the greater diversity of HLA-B alleles in exons 2 and 3 encoding the antigen recognition site is associated with homogeneity in exons 4–8. In contrast, HLA-A alleles that are less diverse in exons 2 and 3 exhibit greater diversity than HLA-B in exons 4–8 (18). From our, albeit small, sampling of gorilla alleles, we see the opposite trends. Excluding Gogo-Okoko, we obtained four very distinctive Gogo-A alleles from four genes isolated, whereas four Gogo-B genes yielded just two distinctive allelic motifs. However, in this small sampling of Gogo-B alleles, there is greater diversity in exon 4 than found in HLA-B. The Gogo-C locus has a lack of diversity similar to Gogo-B, and if the gene organization is like that in humans, then these two loci will be closely linked.

In comparisons of HLA-A, -B, and -C alleles, it has proved easier to track the lineages of HLA-A alleles than either the HLA-B or -C alleles. This is primarily due to the presence of polymorphisms in exons 3–8, particularly exon 4, which are absent in HLA-B. None of these polymorphisms are at functionally important positions of the protein (the antigen recognition site and CD8 binding site); a proportion are silent substitutions and they are thus likely to be selectively neutral.

The HLA-A alleles divide into five families themselves derived from two ancient lineages. One lineage led to the HLA-A2, A10, and A19 families, the other to the HLA-A3 and A9 families. The chimpanzee A alleles (Patr-A) conform more closely to the families defined in the human than do the Gogo-A alleles. Patr-A alleles clearly segregate with the HLA-A3 family, whereas the Gogo-A alleles are a mix of the A10



**Figure 11.** Comparison of the relatedness of Gogo-A and Patr-A alleles to HLA-A alleles. (a) The distribution of differences obtained from pairwise comparisons between HLA-A alleles. (b) The distribution from comparisons between pairs involving one Patr-A allele and one HLA-A allele. (c) Comparisons involving one Gogo-A allele and one HLA-A allele.

and A19 families and appear to have diverged from the human lineage before separation of these two families. This difference does not provide evidence that gorillas are more distantly diverged from humans than chimpanzees because the times at which the families diverged along the 2 lineages are likely to be different.

It is, however, intriguing that the Gogo-A alleles derive from one ancient lineage whereas the Patr-A alleles derive from the other. Thus, although both types of alleles were present in the common ancestor of the humans, gorillas, and chimpanzees, only one type has survived in each of the ape lineages, whereas both have survived in humans. Thus, in comparing A alleles one can see that Gogo-As and Patr-As correspond to two distinct parts of the distribution obtained with the HLA-A alleles (Fig. 11). In considering just this single, highly polymorphic, antigen-presenting locus, humans might be said to be more like both chimpanzees and gorillas than the ape species are to each other.

Address correspondence to Peter Parham, Department of Cell Biology, Sherman Fairchild Building, Stanford University School of Medicine, Stanford, CA 94305. David A. Lawlor's present address is the Department of Immunology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030.

Received for publication 27 June 1991.

## References

- Klein, J., and F. Figueroa. 1986. Evolution of the major histocompatibility complex. *CRC Crit. Rev. Immunol.* 6:295.
- Watkins, D.I., F.S. Hodi, and N.L. Letvin. 1988. A primate species with limited major histocompatibility complex class I polymorphism. *Proc. Natl. Acad. Sci. USA.* 85:7714.
- Orr, H.T., and R. DeMars. 1983. Mapping of class I DNA sequences within the human major histocompatibility complex. *Immunogenetics.* 18:489.
- Lawlor, D.A., J. Zemmour, P.D. Ennis, and P. Parham. 1990. Evidence of class-I MHC genes and proteins: from natural selection to thymic selection. *Annu. Rev. Immunol.* 8:23.
- Parham, P., C.E. Lomen, D.A. Lawlor, J.P. Ways, N. Holmes, H.L. Coppin, R.D. Salter, A.M. Wan, and P.D. Ennis. 1988. Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc. Natl. Acad. Sci. USA.* 85:4005.
- Holmes, N., and P. Parham. 1985. Exon shuffling in vivo can generate novel HLA class I molecules. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2849.
- Sibley, C.G., and J.E. Ahlquist. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J. Mol. Evol.* 20:2.
- Hasegawa, M., H. Kishino, and T. Yano. 1987. Man's place in hominoidea as inferred from molecular clocks of DNA. *J. Mol. Evol.* 26:132.
- Ruvolo, M., T.R. Disotell, M.W. Allard, W.M. Brown, and R.L. Honeycutt. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Natl. Acad. Sci. USA.* 88:1570.
- Lawlor, D.A., F.E. Ward, P.D. Ennis, A.P. Jackson, and P. Parham. 1988. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature (Lond.)* 335:268.
- Mayer, W.E., M. Jonker, D. Klein, P. Ivanyi, G. van Seventer, and J. Klein. 1988. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2765.
- 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.
- Ennis, P.D., J. Zemmour, R.D. Salter, and P. Parham. 1990. Rapid cloning of HLA-A, B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc. Natl. Acad. Sci. USA.* 87:2833.
- Staden, R. 1980. A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* 8:3673.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387.
- Subbiah, S., and S.C. Harrison. 1989. A method for multiple sequence alignment with gaps. *J. Mol. Biol.* 209:539.
- Swofford, D.L. 1990. PAUP: Phylogenetic Analysis Using Parsimony, version 3.0. A computer program distributed by the Illinois Natural History Survey, Champaign, IL.
- Parham, P., D.A. Lawlor, C.E. Lomen, and P.D. Ennis. 1989. Diversity and diversification of HLA-A,B,C alleles. *J. Immunol.* 142:3937.
- Zemmour, J., B.H. Koller, P.D. Ennis, D.E. Geraghty, D.A. Lawlor, H.T. Orr, and P. Parham. 1990. HLA-AR, an inactivated antigen-presenting locus related to HLA-A. Implications for the evolution of the MHC. *J. Immunol.* 144:3619.
- Watkins, D.I., Z.W. Chen, T.L. Garber, A.L. Hughes, and N.L. Letvin. 1991. Segmental exchange between MHC class I genes in a higher primate: recombination in the gorilla between the ancestor of a human non-functional gene and an A locus gene. *Immunogenetics.* 34:185.
- Kato, K., J.A. Trapani, J. Allopenna, B. DuPont, and S.Y. Yang. 1989. Molecular analysis of the serologically defined HLA-Aw19 antigens: a genetically distinct family of HLA-A antigens comprising A29, A31, A32, and Aw33, but probably not A30. *J. Immunol.* 143:3371.
- Ways, J.P., J.B. Rothbard, and P. Parham. 1986. Amino acid residues 56 to 69 of HLA-A2 specify an antigenic determinant shared by HLA-A2 and HLA-B17. *J. Immunol.* 137:217.
- Wan, A.M., P. Ennis, P. Parham, and N. Holmes. 1986. The primary structure of HLA-A32 suggests a region involved in formation of the Bw4/Bw6 epitopes. *J. Immunol.* 137:3671.
- 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 (Lond.)* 329:512.
- Lutz, C.T., D.A. Jensen, J. Schifflbauer, D.K. Didier, B.D. Schwartz, and C.S. Davis. 1990. Multiple mechanisms produce diversity of HLA-C alleles. *Hum. Immunol.* 28:27.
- Pease, L.R., R.M. Horton, J.K. Pullen, and Z. Cai. 1991. Structure and origin of diversity in the major histocompatibility complex. *CRC Crit. Rev. Immunol.* In press.
- Duceman, B.W., D. Ness, R. Rendi, M.J. Chorney, R. Srivastava, D.S. Greenspan, J. Pan, S.M. Weissman, and F.C. Grumet. 1986. HLA-JY 328: mapping studies and expression of a polymorphic HLA class I gene. *Immunogenetics.* 23:90.
- Klein, J. 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum. Immunol.* 19:155.
- Parham, P., and D.A. Lawlor. 1991. Evolution of class I major histocompatibility complex genes and molecules in humans and apes. *Hum. Immunol.* 30:119.
- Lewin, R. 1988. Conflict over DNA clock results. *Science (Wash. DC).* 241:1598.
- Lewin, R. 1988. DNA clock conflict continues. *Science (Wash. DC).* 241:1756.
- Saitou, N., and M. Nei. 1986. The number of nucleotides required to determine the branching order of three species, with

- special reference to the human-chimpanzee-gorilla divergence. *J. Mol. Evol.* 24:189.
33. Hughes, A.L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature (Lond.)*. 335:167.
34. Hedrick, P.W., and G. Thomson. 1983. Evidence for balancing selection at HLA. *Genetics*. 104:449.
35. Townsend, A., and H. Bodmer. 1989. Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol.* 7:601.
36. Ennis, P.D., A.P. Jackson, and P. Parham. 1988. Molecular cloning of bovine class I MHC cDNA. *J. Immunol.* 141:642.
37. Bjorkman, P.J., and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem.* 59:253.