

Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees

Natasja G. de Groot[†], Nel Otting[†], Gaby G. M. Doxiadis[†], Sunita S. Balla-Jhagjhoorsingh[†], Jonathan L. Heeney[†], Jon J. van Rood[‡], Pascal Gagneux[§], and Ronald E. Bontrop^{†¶}

[†]Departments of Immunobiology and Virology, Biomedical Primate Research Centre, P.O. Box 3306, 2280 GH, Rijswijk, The Netherlands; [‡]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, E3-Q, P.O. Box 9600, 2300 RC, Leiden, The Netherlands; and [§]Glycobiology Research and Training Center CMM-East, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0687

Contributed by Jon J. van Rood, July 16, 2002

MHC class I molecules play an essential role in the immune defense against intracellular infections. The hallmark of the MHC is its extensive degree of polymorphism at the population level. However, the present comparison of MHC class I gene intron variation revealed that chimpanzees have experienced a severe repertoire reduction at the orthologues of the *HLA-A*, *-B*, and *-C* loci. The loss of variability predates the (sub)speciation of chimpanzees and did not effect other known gene systems. Therefore the selective sweep in the MHC class I gene may have resulted from a widespread viral infection. Based on the present results and the fact that chimpanzees have a natural resistance to the development of AIDS, we hypothesize that the selective sweep was caused by the chimpanzee-derived simian immunodeficiency virus (SIVcpz), the closest relative of HIV-1, or a closely related retrovirus. Hence, the contemporary chimpanzee populations represent the offspring of AIDS-resistant animals, the survivors of a HIV-like pandemic that took place in the distant past.

The MHC present in most vertebrate species studied encodes two clusters of cell surface proteins. In humans these are designated HLA-A, -B, and -C (class I) and -DP, -DQ, and -DR (class II), respectively. The MHC class I and II gene products play a pivotal role in the induction of adaptive immune responses. MHC class I antigens are expressed on virtually all nucleated cells and bind peptides from intracellular origin (1). Normally they are loaded with self-peptides but in the case of an infection the peptides may originate from viruses (or other intracellular parasites). Cytotoxic T cells (CTL) recognize such MHC class I-peptide complexes as alien, and can trigger the lysis of infected cells. MHC class II molecules, expressed on white blood cells, control antibody production and mediate T-cell help. Apart from self-peptides, they bind peptides usually originating from extracellular pathogens. Polymorphism of the MHC system is mainly confined to the contact residues of the peptide binding site (2). Different MHC molecules select disparate peptides for T-cell activation, and as a consequence particular MHC molecules/alleles are associated with resistance or susceptibility to different infectious diseases. It has been demonstrated that being heterozygous for particular *HLA* alleles may represent an advantage (3). Because of MHC polymorphism, individual variation reduces the chance that one pathogen can sweep through the entire population. In this context, MHC polymorphism could act as an insurance of immunity across a population.

Two species of chimpanzees, which shared an ancestor about 2 million years (myr) ago, have been officially recognized: namely, the common chimpanzee (*Pan troglodytes*, or *Patr*) and the bonobo (*Pan paniscus*, *Papa*) (4). Based on mitochondrial DNA (mtDNA) variation, common chimpanzees have been divided into at least four subspecies designated *P. t. verus* (*Ptv*), *P. t. troglodytes* (*Ptt*), *P. t. schweinfurthii* (*Pts*), and *P. t. vellerosus* (4, 5), although the status of the latter two remains unclear (6).

Humans and chimpanzees display 98.7% similarity at the non-repetitive DNA level and shared an ancestor about 5–6 myr ago (7, 8). Chimpanzees display far more variation in their

mtDNA than humans do (5, 9), and similar findings have been reported for other nuclear genes studied (10, 11). The accepted explanation for these findings is that chimpanzees as a species are older than modern humans and have existed as more subdivided populations, resulting in the accumulation of more variation. In addition, at least one report claims that chimpanzees may have existed at larger effective population sizes than humans (12). Apart from their younger age as a species, humans appear to have undergone multiple population bottlenecks (5, 10, 11).

The MHC system of chimpanzees probably has a genomic structure similar to that of humans; the MHC class I loci are known as *Patr-A*, *-B*, and *-C* (13). For each locus, alleles of shared ancestry can be grouped into lineages, which may predate speciation (14–16). With regard to the number of MHC class I alleles, chimpanzees seem to display at least as much diversity as humans (17). However, for the *A* locus, chimpanzee samples showed only positive typing reactions with particular HLA-A1, -A3, and -A11 alloantisera (13, 14). Subsequent sequencing studies illustrated that chimpanzees only possess orthologues of the *HLA-A1/A3/A11* family, whereas alleles grouping into five other *HLA-A* families appear to be absent (14). In addition, chimpanzees appear to lack the MHC class II equivalents of the *HLA-DRB1*04* and *-DRB1*08* lineages (18). These observations suggest that chimpanzees may have lost certain MHC lineages during evolution. The absence of these particular lineages can be safely considered to represent a loss in chimpanzees rather than a recent gain in humans because of the trans-species mode of evolution of MHC lineages (19).

MHC class I and II sequences encode gene products that were shown to be under frequency-dependent/diversifying selection (20–22). This, and the imbalance in sample size between the number of humans and chimpanzees analyzed, can hamper an accurate interpretation of the data with regard to a loss of alleles/lineages because of disease susceptibility. Consequently, we studied the potential influence of negative/purifying selection operating on the MHC class I *A*, *B*, and *C* loci by comparing intron variation in humans and chimpanzees, because introns are known to evolve in a neutral fashion (23).

Materials and Methods

Animals. The Biomedical Primate Research Centre chimpanzee colony started with 35 founder animals originating from Sierra Leone and belonging to the subspecies *P. t. verus* (West Africa). Included in the study are five animals of the *P. t. troglodytes* (Central Africa) and four animals of the *P. t. schweinfurthii* (East Africa) subspecies. In addition, three animals of the *P. t. verus* subspecies originating from other colonies were studied. The

Abbreviations: myr, million years; *Patr*, *Pan troglodytes*; SIVcpz, chimpanzee-derived simian immunodeficiency virus.

Data deposition: The sequences reported in this paper have been deposited in the EMBL database (accession nos. AJ313335–AJ313359).

[¶]To whom reprint requests should be addressed. E-mail: bontrop@bprc.nl.

animals are characterized on the molecular level for MHC class I and II gene polymorphisms (13, 18). Their offspring have been pedigreed based on segregation of serological specificities (*Patr-A* and *-B*) and molecular-defined *Patr* class II gene polymorphisms.

DNA Amplification and Sequencing. Genomic DNA (gDNA), obtained from Epstein–Barr virus transformed B-cell lines, was used to amplify a ± 950 -bp fragment [containing complete exon 2 (270 bp), intron 2 (241 bp), and exon 3 (276 bp), and partly intron 1 (± 95 bp) and intron 3 (± 60 bp)], using MHC locus-specific primers: 5AIn/3AIn for the *Mhc-A* locus, 5BIn/3BIn for the *-B* locus, and 5CIn/3CIn for the *-C* locus (24). PCR (25 μ l) contained gDNA (0.5 μ g), 0.2 μ M of each primer, 1.5 mM MgCl₂ (for the *B* locus occasionally a concentration of 1.75 mM MgCl₂ is necessary), 0.2 mM of each deoxyribonucleoside triphosphate (dNTP), and 0.5 units of *Taq* polymerase. A total of 33 cycles were run, each cycle consisting of 30 s at 95°C, 50 s at 65°C, and 30 s at 72°C with a final amplification of 8 min at 72°C. The PCR reactions were purified using the QIAquick Gel Extraction Kit (Qiagen). Purified PCR products were sequenced directly on the ABI 310 automatic sequencer (Applied Biosystems) by using the above-mentioned locus-specific primers. The products were sequenced from the 5' and 3' ends. Cycle sequencing reactions were carried out with ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) according to the manufacturer's instructions. At least two independent PCR reactions were performed and/or intron 2 sequences were confirmed by their presence in different MHC typed animals.

Phylogenetic Analysis. The UPGMA phylogenetic tree of intron 2 sequences was constructed by using the pairwise genetic distances method calculated by using the Jukes–Cantor correction for multiple hits (25) and rooted by the midpoint method. This allows translation of genetic distances into divergence time (26). For the construction, the computer program PAUP* v.4.0B8 for Macintosh was used (44). The same program was used to define bootstrap values based on 1,000 resamplings. Trees constructed using neighbor joining and maximum likelihood (quartet puzzling) gave very similar topologies in the sense that chimpanzee sequences were only found on relatively few deep clades.

Neutrality Test. Tajima's D test (27) and Fu's Fs test (28) were performed using ARLEQUIN v.2.0, a software for population genetics data analysis (Genetics and Biometry Laboratory, University of Geneva; <http://anthro.unige.ch/arlequin>).

Results and Discussion

Comparative Analysis of MHC Class I Intron 2 Sequence Variation in Humans and Chimpanzees. Intron 2 (241 bp), situated between the polymorphic exons 2 and 3, is a valuable gene segment for studying the evolution of MHC class I genes. Various *HLA* and *Patr* class I alleles, from different lineages, have been analyzed for their neighboring intron 2 sequence. An overview of the number of distinct sequences identified is provided in Table 1.

Fig. 1, summarizing the *HLA-* and *Patr-A*, *-B*, and *-C* intron 2 sequences, depicts the polymorphic nucleotide positions for the different alleles. For the *HLA-A* locus most intron 2 sequences appear to be specific for a lineage (Fig. 1A). The variation in the *HLA-A* exon 2 and 3 sequences evolved, however, mainly by point mutations and thus reflects diversifying selection in contrast to the apparently neutral evolution operating on the introns. Although to a very minor extent, sharing of identical intron 2 sequences between different *HLA-A* lineages evidences recombination (Table 2). Compared with those of humans the chimpanzee *A*-locus, intron 2 sequences possess fewer unique nucleotide substitutions and are far less heterogeneous (Fig.

Table 1. Number of MHC class I coding alleles and intron 2 sequences in humans and chimpanzees

Locus	No. MHC class I coding alleles	N*	No. intron 2 sequences
<i>HLA-</i>			
<i>A</i>	243	40	15
<i>B</i>	478	33	23
<i>Cw</i>	115	15	12
<i>Patr-</i>			
<i>A</i>	26	18	8
<i>B</i>	43	16	11
<i>C</i>	21	11	6

S. G. E. Marsh and J. Robinson donated the annotated *HLA-A* intron 2 data. The other *HLA* intron 2 sequences were collected from various databases, whereas the chimpanzee intron 2 data are from this study.

*Number of MHC alleles that are studied for their neighboring intron 2 sequence.

1A). Only ten species-unique nucleotide substitutions are observed in *Patr-A* versus 23 in the *HLA-A* intron 2 sequences.

The lineage specificity for intron 2 sequences at the *HLA-B* locus is less strict than for the *HLA-A* locus (Fig. 1B). Hence, recombination appears to affect the order of intron–exon sequences, as found for *HLA-B* exons (16, 31, 32). Only members of *HLA-B*08*, *-B*27*, and *-B*35* seem to have lineage unique intron 2 sequences. The *Patr-B* intron 2 sequences, on the other hand, can be divided into two homogeneous clusters, one of which is characterized by a three-nucleotide deletion (Fig. 1B). A comparison demonstrates that only ten *Patr-B*-unique nucleotide substitutions are found, whereas 25 *HLA-B*-specific substitutions are reported (Fig. 1B). Twelve unique nucleotide substitutions are left when the *HLA-B*7301* sequence is ignored. However, this still indicates a greater diversity in advantage of the human population. The data show that recombination promotes diversification at the *Patr-B* locus to a lesser extent than in *HLA-B* (Table 2). If a selective sweep occurred in chimpanzees, particular *Patr-B* intron 2 lineages were lost because of the negative selection (Table 3). Hence, as compared with humans the reservoir of intron 2 sequences in chimpanzees is relatively small and as a consequence recombination should be observed less prominently.

At *HLA-Cw*, only a limited set of intron 2 data have been reported (Fig. 1C). The *HLA-Cw*0303* and *-Cw*0304* alleles share an identical intron 2 sequence. Furthermore, the intron 2 sequences of *HLA-Cw*0701* and *-Cw*0702* differ by only one nucleotide, whereas intron 2 of *HLA-Cw*0701* also contains an insert of five nucleotides (Fig. 1C). The present data suggest a situation similar to that observed for *HLA-A* intron 2 data: namely, that every lineage has its own characteristic intron 2 sequence (Table 2). The *Patr-C* intron 2 sequences seem to constitute two main clusters, which allow a further division into four lineages (Fig. 1C). The intron 2 sequences of *Patr-C*0203*, *-C*1303*, *-C*02/*13*, and *-C*0401* have large parts in common indicating that they arose from one ancestor. The *Patr-C*02/*13* sequences could have resulted from a recombination between the intron 2 sequences of *Patr-C*0203*, *-C*1303*, and *-C*0401*. Twenty unique *HLA-Cw* substitutions are found in intron 2 versus 16 *Patr-C* specific nucleotides.

The current findings reveal that the *HLA-A*, *-B*, and *-C* intron 2 sequences accumulated or maintained more variation than their chimpanzee counterparts. It is generally accepted that introns of the same gene systems in humans and great apes evolve under identical or nearly identical neutral conditions (23). This argues against the possibility that humans accumulated more variation than chimpanzees over a relatively short time span. Our observation of low MHC class I intron 2 variation in

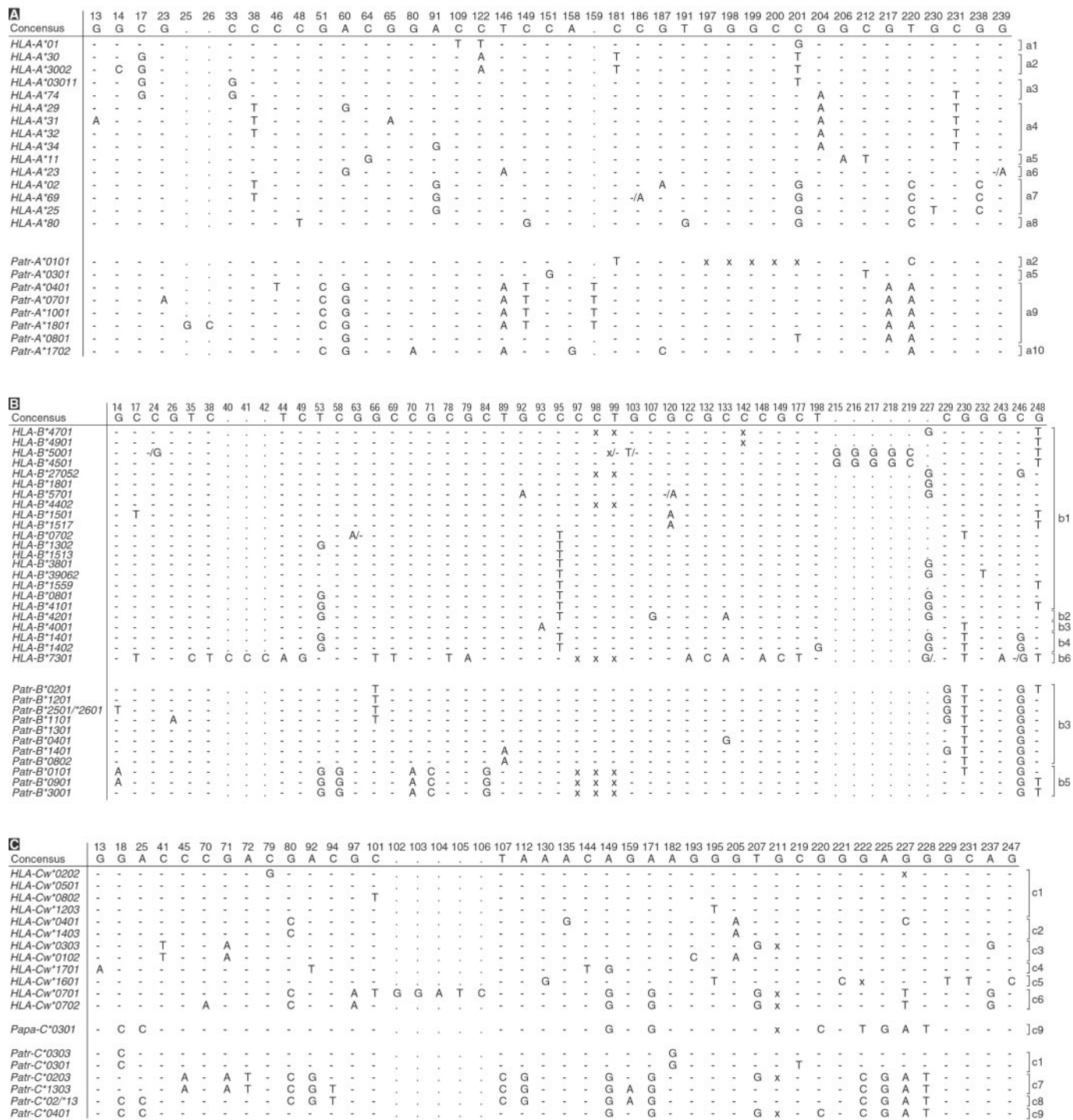


Fig. 1. Polymorphic nucleotide positions in *HLA* and *Patr* intron 2 sequences. Identity to the consensus sequence (depicted at the top) is indicated by dashes. Substitutions and inserts are depicted by the conventional one-letter code, deletions are marked by "x". For instance, "-/A" indicates that differences in a particular sequence have been reported in the literature. When multiple alleles of a lineage share an intron 2 sequence, only a representative sequence is shown (mentioning only the first two digits in the name). In some cases two allele designations are mentioned when it is not clear which allele couples to this particular intron 2 sequence. The brackets show the division of the alleles into trans-species lineages based on phylogenetic analysis (Fig. 2). (A) *Mhc-A* locus data. (B) *Mhc-B* locus data. The intron 2 sequences of *Patr-B*0101* and *-B*0401* are identical to published sequences ChB.Cr12 and ChB.Ch18, respectively (29). (C) *Mhc-C* locus data. The figure shows also an intron 2 sequence obtained from a bonobo (*Papa-C*0301*; ref. 30).

chimpanzees versus humans sharply contrasts the situation documented for other neutrally evolving genomic segments of these two species (5, 9–11). As such it provides evidence that ancestral chimpanzee populations have experienced a selective sweep leading to the loss of particular MHC lineages. The reduced

intron 2 variation in chimpanzees versus humans is reflected not only by a limited amount of nucleotide variation but also by a low number of intron 2 sequences. One could argue that our sample size is too small. However, all chimpanzee samples added (8 *Ptv*, 2 *Pts*, 1 *Ptt*) did not result in the detection of novel intron 2

Table 2. *HLA* and *Patr* class I alleles that share an identical intron 2 sequence

<i>Mhc</i> locus	Name intron 2 sequence	Coupled MHC class I coding allele
A	<i>HLA-A*23</i>	<i>HLA-A*2301</i> , <i>-A*24</i>
	<i>HLA-A*25</i>	<i>HLA-A*25</i> , <i>-A*2601</i> , <i>-A*6601</i> , <i>-A*68</i>
	<i>HLA-A*31</i>	<i>HLA-A*31</i> , <i>-A*33</i>
	<i>Patr-A*0101</i>	<i>Patr-A*0101</i> , <i>-A*0201</i> , <i>-A*0601</i> , <i>-A*0602</i> , <i>-A*1101</i>
	<i>Patr-A*0301</i>	<i>Patr-A*0301</i> , <i>-A*0302</i>
	<i>Patr-A*0401</i>	<i>Patr-A*0401</i> , <i>-A*0501</i> , <i>-A*0901</i> , <i>-A*1201</i> , <i>-A*1301</i> , <i>-A*1401</i>
B	<i>HLA-B*1501</i>	<i>HLA-B*1501</i> , <i>-B*4601</i>
	<i>HLA-B*1513</i>	<i>HLA-B*1513</i> , <i>-B*35</i> , <i>-B*5301</i> , <i>-B*5401</i> , <i>-B*5601</i> , <i>-B*5801</i>
	<i>HLA-B*1801</i>	<i>HLA-B*1801</i> , <i>-B*3701</i>
	<i>HLA-B*27052</i>	<i>HLA-B*27</i> , <i>-B*4002/B*4003</i>
	<i>HLA-B*3801</i>	<i>HLA-B*3801</i> , <i>-B*4102</i>
	<i>Patr-B*0101</i>	<i>Patr-B*0101</i> , <i>-B*0301</i>
	<i>Patr-B*0401</i>	<i>Patr-B*0401</i> , <i>-B*16011</i>
	<i>Patr-B*1201</i>	<i>Patr-B*1201</i> , <i>-B*1202</i> , <i>-B*2402</i>
	<i>Patr-B*1401</i>	<i>Patr-B*1401</i> , <i>-B*2901</i>
C	<i>HLA-Cw*0501</i>	<i>HLA-Cw*0501</i> , <i>-Cw*0602</i>
	<i>HLA-Cw*1403</i>	<i>HLA-Cw*1403</i> , <i>-Cw*1801</i>
	<i>Patr-C*0303</i>	<i>Patr-C*0303</i> , <i>-C*0501</i> , <i>-C*0502</i> , <i>-C*0601</i> , <i>-C*1101</i>
	<i>Patr-C*0401</i>	<i>Patr-C*0401</i> , <i>-C*0901</i>

The table lists, for both species, the intron 2 sequences (second column) with the coupled MHC class I coding alleles (third column). Only the first two digits of the lineage name are mentioned (for example, *HLA-A*25*) when alleles of a lineage share the same intron 2 sequence.

sequences. Moreover, in a comparison between 25 MHC-typed chimpanzees of the Biomedical Primate Research Centre colony and 25 MHC-typed humans of the Dutch population, χ^2 statistics shows that the intron 2 variation found in humans is 2.56 times higher [confidence interval (CI) 95% is 0.87–7.55, $P = 0.07$] for the A locus and 2.64 times (CI 95% is 1.20–5.82, $P = 0.01$) for the B locus. Furthermore, MHC class I intron 3 (586 bp) sequence analysis for particular *HLA-Cw* and *Patr-C* alleles indicates that the reduced sequence variation is not intron 2-specific and extends to other gene segments (data not shown). The diversity encountered in the chimpanzee mtDNA and in other genetic systems illustrates, however, that a relatively large population of chimpanzees survived the selective sweep.

Phylogenetic Analysis of *HLA* and *Patr* Class I Intron 2 Sequences. Humans and chimpanzees shared a common ancestor ≈ 5 –6 myr ago. The phylogenetic tree illustrates that at least ten *Mhc-A* (a1–a10), six *-B* (b1–b6), and nine *-C* (c1–c9) lineages predate the speciation of humans and chimpanzees (Fig. 2). Barring convergent evolution, the existence of alleles of one species with closest relatives in the other species suggests the existence of trans-species lineages (as is illustrated, for instance, by the lineage a2 in Fig. 2).

In a test for neutrality (Tajima's *D*; ref. 27), none of the chimpanzee intron 2 sequences considered here showed departures from neutrality as is indicated by *D* statistics and their *P* values (Table 4). We cannot rule out a more recent diversifying selection at the *HLA-B* locus based on *D* value, and the high negative value of Fu's *F_s* test (28) suggests a recent demographic expansion. This should not surprise us, considering that the *HLA-B* locus is the most polymorphic locus known, with over 400 alleles.

For the *Mhc-A* locus, human alleles represent eight different lineages, whereas chimpanzee alleles represent four lineages

Table 3. Distribution of MHC class I intron 2 lineages and their members in humans (*HLA*) and chimpanzees (*Patr*)

<i>Mhc</i> -locus	Intron 2 lineage	Species	
		<i>HLA</i>	<i>Patr</i>
A	a1	+	–
	a2	+	+
	a3	+	–
	a4	+	–
	a5	+	+
	a6	+	–
	a7	+	–
	a8	+	–
	a9	–	+
	a10	–	+
B	b1	+	–
	b2	+	–
	b3	+	+
	b4	+	–
	b5	–	+
	b6	+	–
C	c1	+	+
	c2	+	–
	c3	+	–
	c4	+	–
	c5	+	–
	c6	+	–
	c7	–	+
	c8	–	+
	c9	–	+

+, Presence of the lineage; –, absence of the lineage. The division into the trans-species intron 2 lineages, a1–a10, b1–b6, and c1–c9, is based on Fig. 2.

(Table 3). Only the a2 and a5 lineages are shared. These results suggest that chimpanzees have lost the representatives of the a1, a3, a4, a6, a7, and a8 lineages, whereas humans may have lost only the evolutionary equivalents of the a9 and a10 lineages. The alternative interpretation is that modern humans accumulated rapid variation in the MHC class I intron sequences. This reasoning is in conflict with the old age of many human *HLA-A* intron 2 lineages, as well as with the neutral theory of evolution.

At the *Patr-A* locus, eight distinct intron 2 sequences have been detected clustering into four different trans-species lineages. These intron lineages are coupled to the highly homogeneous *Patr-A* coding sequences, which all cluster into the *HLA-A1/A3/A11* family. Only the a5 lineage contains a *Patr-A* intron linked to the family of *HLA-A1/A3/A11* exons, whereas other *Patr* introns cluster into other lineages (Fig. 2). This indicates that in the past different chimpanzee introns may have been coupled to exons clustering in different lineages because of recombination. Because of negative selection many exons (except equivalents of the *HLA-A1/A3/A11* family) and their corresponding introns were lost during chimpanzee evolution. For that reason some of these *Patr-A* locus introns, which do not show similarities toward *HLA-A* introns linked to the *HLA-A1/A3/A11* family, may represent relics from ancient recombination events.

At the *Mhc-B* locus, humans possess alleles from five lineages and chimpanzees only from two lineages, whereas at the *Mhc-C* locus humans and chimpanzees have alleles from six and four lineages, respectively (Table 3).

Thus, on average, chimpanzees have lost more intron 2 lineages than have humans during ≈ 5 myr of evolution. Moreover, chimpanzees appear to show considerably less intron 2 sequence variation than their human equivalents, as is reflected in the phylogenetic tree by the limited number of distinct clades

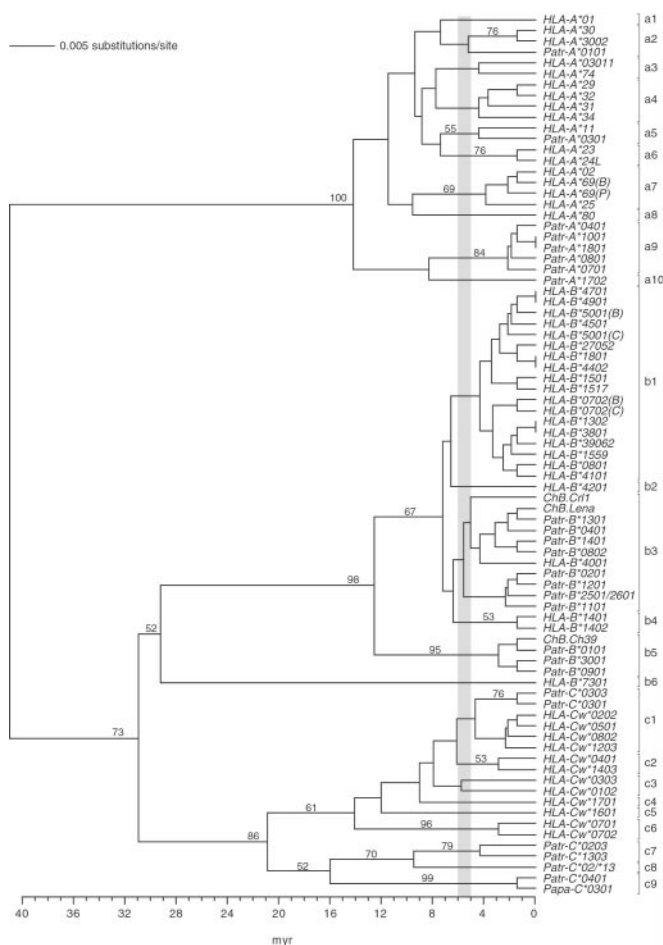


Fig. 2. UPGMA phylogenetic tree of human and chimpanzee MHC class I intron 2 sequences. The divergence time was estimated based on a substitution rate for introns of 1.4×10^{-9} per site per year (33). The bar between 5 and 6 myr highlights the period when humans and chimpanzees shared a common ancestor. The brackets indicate the division in trans-species lineages. The relevant bootstrap values are indicated. The sequences of ChB.Crl1, ChB.Ch39, and ChB.Lena were published (29). In some cases two allele designations are mentioned when it is not clear which allele couples to the particular intron 2 sequence. B, C, or P indicates that for these alleles differences are reported in the literature.

and/or clades with short branch lengths (Fig. 2). These observations are also in agreement with the fact that chimpanzees show less variability at the coding sequences for the classical MHC class I molecules.

Time Estimate for a Selective Sweep. The subspeciation of chimpanzees occurred ≈ 1.5 myr ago (4). Comparative analyses show that chimpanzee subspecies share identical MHC class I intron 2 alleles, indicating that some of these sequences have been genetically stable over a relatively long time span. Overall, chimpanzee subspecies share MHC class I intron 2 lineages, suggesting that the repertoire reduction took place before the subspeciation.

At present, only limited data exist on bonobo MHC class I intron (30) and exon (14) sequences. The consensus view is that bonobos also only have alleles that cluster into the *HLA-A1/A3/A11* family. This would indicate that the selective sweep also predates the speciation of common chimpanzees and bonobos.

The MHC class I intron 2 sequence variation present in chimpanzees appears to be of relatively recent origin. For

Table 4. Tajima's D values and Fu's Fs values for human and chimpanzee intron 2 sequences

	D	P value	Fs
<i>HLA-A</i>	-0.69356	0.26621	-12.25739
<i>Patr-A</i>	-0.66281	0.29455	-3.89288
<i>HLA-B</i>	-1.69664	0.03586	-34028234
<i>Patr-B</i>	0.30608	-0.37849	-10.15467
<i>HLA-Cw</i>	-0.87723	0.21324	-6.91030
<i>Patr-C</i>	1.31301	-0.12144	-1.46290

The P value is calculated for Tajima's D.

chimpanzees, the common ancestral intron 2 alleles in the a9, b3, b5, and c1 lineages have ages around 1.5–3 myr (Fig. 2).

Taken together, the selective sweep causing the MHC repertoire reduction in chimpanzees must have occurred before the (sub)speciation of chimpanzees. Considering the age of the intron 2 lineages, the selective sweep is dated to have happened ≈ 2 –3 myr ago.

The Cause of the MHC Class I Repertoire Reduction: A Hypothesis.

MHC class I molecules play a critical role in the immune defense against intracellular infections caused, for instance, by viruses (1). Therefore, the MHC class I repertoire reduction in chimpanzees may have resulted from a widespread viral infection in the ancestral populations of the contemporary chimpanzee species. Although the effect is most prominent in the MHC class I region, chimpanzees have apparently lost particular MHC class II lineages as well (18). This is not surprising, as many viruses have both dominant intra- and extracellular stages of infection and MHC class II-mediated antibody responses are therefore also an important correlate of protection. On the other hand, particular MHC class II lineages may also have been lost because of preferential physical linkage (close proximity on the chromosome) to MHC class I alleles that came under negative selection.

The question arises: Which pathogen may be held responsible for the selective sweep in the MHC class I repertoire of chimpanzees? Humans and chimpanzees are the only known species susceptible to infection with pathogens like HIV, HCV, and *Plasmodium falciparum*. However, both species may show marked differences in pathology after infection (34–36).

Natural infections with SIVcpz, the closest relative of HIV-1, have been documented in at least six chimpanzees (37) and one free-ranging wild chimpanzee (38). Furthermore, a recent study estimated the zoonotic event with SIVcpz/HIV-1, which gave rise to the human AIDS epidemic, to have taken place approximately 70 years ago (39). In this context, the ancestor of SIVcpz/HIV-1 could be considered as a prime candidate for the selective sweep in the MHC class I repertoire of chimpanzees. In the past, worldwide, approximately 150 chimpanzees were infected with various HIV-1 strains, but only one animal was diagnosed with symptoms of AIDS (40). This particular animal was co-infected with different HIV-1 isolates, and the virus isolated at the time of disease was a recombinant that apparently escaped existing immune responses (41). The relative resistance of HIV-1-infected chimpanzees to the development of AIDS may be the consequence of an effective immune response controlled, at least in part, by the present set of MHC class I molecules, which are the result of positive selection. We have recently demonstrated that at least some chimpanzee MHC class I-restricted immune responses target conserved epitopes of the HIV-1 virus (34), resulting in effective control of infection. These *Patr* alleles are characterized by relatively high frequency numbers (13). Identical viral epitopes are recognized by human long-term nonprogressors in the context of particular HLA class I molecules associated with resistance (42). However, humans

and chimpanzees recognize such epitopes in the context of MHC class I molecules that group into distinct lineages (34), thus illustrating that the quality of MHC molecules to bind particular peptides may determine whether an individual is susceptible or resistant to a disease.

Some orthologues of the *HLA-A1/A3/A11* lineage in chimpanzees have *HLA-A2* binding motifs (43). This example illustrates that one has to be careful to extrapolate structural chimpanzee data into consequences for the human situation with regard to the function of MHC class I molecules. For that reason, peptide binding studies have been initiated to determine whether chimpanzee MHC class I molecules indeed preferentially target epitopes mapping to conserved regions of SIVcpz/HIV-1.

Santiago *et al.* (38) suggested that the geographic isolation of *P. t. verus* predated the infection by the SIVcpz progenitor. The presently described MHC class I repertoire reduction is observed in all chimpanzee subspecies, as well as the resistance to devel-

oping AIDS. We therefore put forward the hypothesis that ancestors of today's chimpanzee populations went through a pandemic caused by SIVcpz or a related ancestral retrovirus. As a consequence, contemporary chimpanzee populations have modified MHC repertoires, partially reduced, but able to cope with their natural environment and with retroviral infections such as SIVcpz/HIV-1. The fact that SIVcpz is not readily detectable in wild-ranging chimpanzees may be explained, in part, by an effective cytotoxic T cell (CTL) immune response in resistant animals that eventually contained and controlled virus spread in the population.

We thank W. F. Bodmer and P. Parham for positive criticism and support, D. Devine for editing the manuscript, and H. van Westbroek for the artwork. We also thank C. Walker for providing the DNA of additional chimpanzees, S. G. E. Marsh for sharing HLA intron data, and E. Remarque for statistical help. This study was financed in part by the European Union Project, IMGT-QLG2-CT-2000-01287.

- Parham, P. & Ohta, T. (1996) *Science* **272**, 67–74.
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329**, 512–518.
- Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K. & O'Brien, S. J. (1999) *Science* **283**, 748–752.
- Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J. & Woodruff, D. S. (1994) *Science* **265**, 1193–1201.
- Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P. A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O. A. & Woodruff, D. S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 5077–5082.
- Gagneux, P., Gonder, M. K., Goldberg, T. L. & Morin, P. A. (2001) *Philos. Trans. R. Soc. London B* **356**, 889–897.
- Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T. D., Itoh, T., Tsai, S. F., Park, H. S., Yaspo, M. L., Lehrach, H., Chen, Z., *et al.* (2002) *Science* **295**, 131–134.
- Sibley, C. G. & Ahlquist, J. E. (1987) *J. Mol. Evol.* **26**, 99–121.
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000) *Nature (London)* **408**, 708–713.
- Kaessmann, H., Wiebe, V. & Pääbo, S. (1999) *Science* **286**, 1159–1162.
- Zhao, Z., Jin, L., Fu, Y., Ramsay, M., Jenkins, T., Leskinen, E., Pamilo, P., Trexler, M., Patthy, L., Jorde, L. B., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11354–11358.
- Chen, F. C. & Li, W. H. (2001) *Am. J. Hum. Genet.* **68**, 444–456.
- de Groot, N. G., Otting, N., Argüello, R., Watkins, D. I., Doxiadis, G. G. M., Madrigal, J. A. & Bontrop, R. E. (2000) *Immunogenetics* **51**, 398–409.
- McAdam, S. N., Boyson, J. E., Liu, X., Garber, T. L., Hughes, A. L., Bontrop, R. E. & Watkins, D. I. (1995) *J. Immunol.* **154**, 6421–6429.
- Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. & Parham, P. (1988) *Nature (London)* **335**, 268–271.
- McAdam, S. N., Boyson, J. E., Liu, X., Garber, T. L., Hughes, A. L., Bontrop, R. E. & Watkins, D. I. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5893–5897.
- Adams, E. J., Cooper, S., Thomson, G. & Parham, P. (2000) *Immunogenetics* **51**, 410–424.
- Bontrop, R. E., Otting, N., de Groot, N. G. & Doxiadis, G. G. M. (1999) *Immunol. Rev.* **167**, 339–350.
- Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G. & Klein, J. (1988) *EMBO J.* **7**, 2765–2774.
- Bodmer, W. F. (1972) *Nature (London)* **237**, 139–145.
- Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
- Hughes, A. L. & Nei, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 958–962.
- Kimura, M. (1968) *Nature (London)* **217**, 624–626.
- Cereb, N., Maye, P., Lee, S., Kong, Y. & Yang, S. Y. (1995) *Tissue Antigens* **45**, 1–11.
- Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–32.
- Bergström, T. F., Josefsson, A., Erlich, H. A. & Gyllensten, U. (1998) *Nat. Genet.* **18**, 237–242.
- Tajima, F. (1989) *Genetics* **123**, 585–595.
- Fu, Y. (1997) *Genetics* **147**, 915–925.
- Cereb, N., Kong, Y., Lee, S., Maye, P. & Yang, S. Y. (1996) *Tissue Antigens* **47**, 498–511.
- Cooper, S., Adams, E. J., Wells, R. S., Walker, C. M. & Parham, P. A. (1998) *Immunogenetics* **47**, 212–217.
- Belich, M., Madrigal, J. A., Hildebrand, W. H., Zemmour, J., Williams, R. C., Petzl-Erler, M. L. & Parham, P. (1992) *Nature (London)* **357**, 326–329.
- Watkins, D. I., Chen, Z. W., Hughes, A. L., Evans, M. G., Tedder, T. F. & Letvin, N. L. (1990) *Nature (London)* **346**, 60–63.
- Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. & Hewett-Emmett, D. (1996) *Mol. Phylogenet. Evol.* **5**, 182–187.
- Balla-Jhaghihoorsingh, S. S., Koopman, G., Mooij, P., Haaksma, A. G. M., Teeuwssen, V. J. P., Bontrop, R. E. & Heeney, J. L. (1999) *J. Immunol.* **162**, 2308–2314.
- Cooper, S., Erickson, A. L., Adams, E. J., Kansopon, J., Weiner, A. J., Chien, D. Y., Houghton, M., Parham, P. & Walker, C. M. (1999) *Immunity* **10**, 439–449.
- Daubersies, P., Thomas, A. W., Millet, P., Brahimi, K., Langermans, A. M. J., Ollomo, B., Ben Mohamed, L., Slierendregt, B., Eling, W., van Belkum, A., *et al.* (2000) *Nat. Med.* **6**, 1258–1263.
- Hahn, B. H., Shaw, G. M., De Cock, K. & Sharp, P. M. (2000) *Science* **287**, 607–614.
- Santiago, M. L., Rodenburg, C. M., Kamenya, S., Bibollet-Ruche, F., Gao, F., Bailes, E., Meleth, S., Soong, S. J., Kilby, J. M., Moldoveanu, Z., *et al.* (2002) *Science* **295**, 465.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. (2000) *Science* **288**, 1789–1796.
- Novembre, F. J., Saucier, M., Anderson, D. C., Klumpp, S. A., O'Neil, S. P., Brown, C. R., Jr., Hart, C. E., Guenther, P. C., Swenson, R. B. & McClure, H. M. (1997) *J. Virol.* **71**, 4086–4091.
- Mwaengo, D. M. & Novembre, F. J. (1998) *J. Virol.* **72**, 8976–8987.
- Kaslow, R. A., Carrington, M., Apple, R., Park, L., Munoz, A., Saah, A. J., Goedert, J. J., Winkler, C., O'Brien, S. J., Rinaldo, C., *et al.* (1996) *Nat. Med.* **2**, 405–411.
- Bertoni, R., Sette, A., Sidney, J., Guidotti, L. G., Shapiro, M., Purcell, R. & Chisari, F. V. (1998) *J. Immunol.* **161**, 4447–4455.
- Swofford, D. L. (1992) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods) (Sinauer, Sunderland, MA), Version 4.