A Selective Sweep in Chimpanzees: Is SIV the Culprit?

A Selective Sweep in Chimpanzees: Is SIV the Culprit?

Een repertoire reductie in chimpansees: Is SIV de oorzaak?

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 22 september 2010 des middags te 12.45 uur

door

Natasja Gabry de Groot

geboren op 18 april 1972 te Den Haag

Promotor: Prof.dr. R.E. Bontrop

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van het B.P.R.C.

Het leven gaat niet over rozen,
toch is iedere dag van het leven deze bijzondere bloem meer dan waard.

Aan mijn ouders
Voor Patrick, Mirte en Tycho

Beoordelingscommissie:

Prof. dr. R.J. de Boer
Prof. dr. W. van Eden
Prof. dr. P. Hogeweg
Prof. dr. F. Miedema
Prof. dr. D.I. Watkins

The research described in this thesis was performed at the Department of Comparative Genetics and Refinement of the Biomedical Primate Research Centre, Rijswijk, The Netherlands.

Illustrations	Henk van Westbroek
Cover	Nico de Groot
Lay out	Henk van Westbroek/Natasja de Groot
Printed by	Gildeprint Drukkerijen BV, Enschede
ISBN Number	978-90-393-5372-1

Table of contents

Chapter 1

Introduction

General introduction

The major histocompatibility complex (MHC) plays a key role in the immune responses of all contemporary living vertebrate species; it is studied most thoroughly in humans, and is of interest to various researchers active in different fields of science. From an evolutionary perspective, humans and chimpanzees are each other's closest living relatives. Next to humans, chimpanzees are susceptible to infections with human immunodeficiency virus type 1 (HIV-1), various hepatitis viruses, and *Plasmodium falciparum*, and are therefore considered to be an important resource for biomedical research. Studies in the twentieth century's mid-1990s hinted that in comparison to humans, chimpanzees may have a reduced repertoire at one of the classical MHC class I loci. However, a thorough population study on chimpanzees was lacking. Together with the knowledge that chimpanzees can be infected with HIV-1 but do not develop acquired immunodeficiency syndrome (AIDS), the aforementioned observations led to genetic and cellular studies on the pedigreed West African chimpanzee colony housed at the Biomedical Primate Research Centre (BPRC) to sort out the nature of this natural resistance.

Taxonomy of primate species

Taxonomy is defined as the theoretical study of the classification of organisms, and it facilitates the ordering of organisms into a group on the basis of shared features. At present, approximately 350 different primate species are recognized. These can be divided into the suborder Strepsirrhini ("wet-nosed" primates) and Haplorrhini ("dry-nosed" primates) (1). The Strepsirrhini, comprising the non-tarsier prosimians, are of a more primitive morphology than the Haplorrhini (1, 2). The suborder Haplorrhini is divided into the infraorder Tarsiiformes (comprising the tarsiers) and Simiiformes (comprising the monkeys and apes). The latter is further subdivided into the infraorders Platyrrhini (New World monkeys) and Catarrhini (Old World monkeys and apes). The Platyrrhini comprises the families Cebidae, Aotidae, Pitheciidae and Atelidae, whereas the Catarrhini are grouped into the superfamily Cercopithecoidea, which includes the family Cercopithecidae. Apes and humans are placed into the superfamily Hominoidea, which in turn includes the families Hylobatidae (gibbons and "lesser apes") and Hominidae (humans and great apes) (Fig. 1). The chimpanzee (*Pan troglodytes*) and pygmy chimpanzee/bonobo (*Pan paniscus*) belong to the Hominidae, and share a common ancestor that lived about 2 million years ago (3). The chimpanzee is regarded as the closest living relative of humans; both species display approximately 98.7% similarity at the non-repetitive DNA level, and they share a common ancestor living approximately 5-6 million years ago (4, 5). Based on mitochondrial DNA (mtDNA) sequence diversity, and on geographic distribution three different subspecies of chimpanzees are recognized (Fig. 2) (3, 6). There may be a fourth subspecies, *P.t.vellerosus*, which has its habitat in Nigeria and

parts of Cameroon (7). Animals of the colony studied in this thesis mainly originate from West Africa (Sierra Leone), and belong to the subspecies *P.t.verus*.

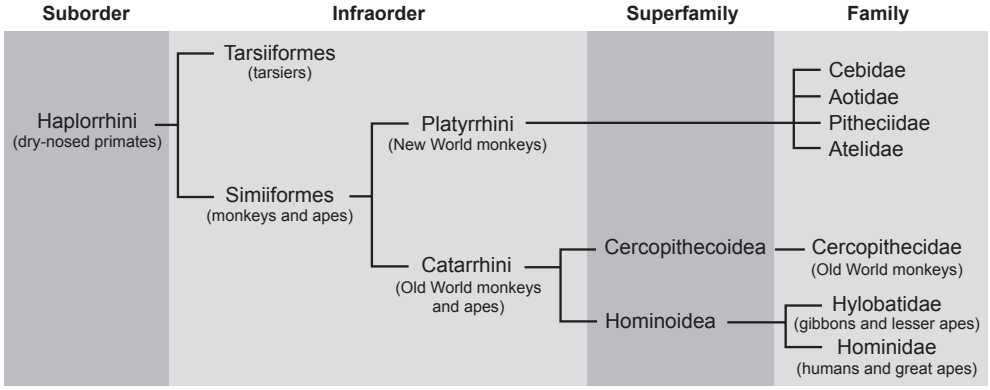


Fig. 1. Overview of the taxonomy of the dry-nosed primates.

The major histocompatibility complex

Historical aspects

The MHC was discovered in 1936 by J.B.S. Haldane's student P.A. Gorer. Haldane was inspired by the work on tumor transplantation, and put forward ideas about why tumors did not grow in certain foreign hosts. He proposed that tumor resistance factors were similar to blood group antigens (described by P. Ehrlich and J. Morgenroth in 1900), and that the failure of the transplants to grow was just like the destruction of incompatible blood cells following transfusion. Most work on tumor transplantation was done initially in mice, and therefore Gorer first had to define blood group antigens in this species. He succeeded in obtaining an antiserum that was able to distinguish three antigens. Accordingly, he wished to answer the question as to whether these antigens were related to the tumor resistance factors. By transplanting tumors into different mouse strains and their hybrids, he discovered, on the one hand, that all mice that lacked antigen II rejected the tumor, but, on the other hand, animals that possessed antigen II resisted the tumor. He also observed that antigen II was present on normal and malignant tissue. Based on his observations, Gorer posited the theory that the tumor resistance factors were antigens that were shared by malignant and

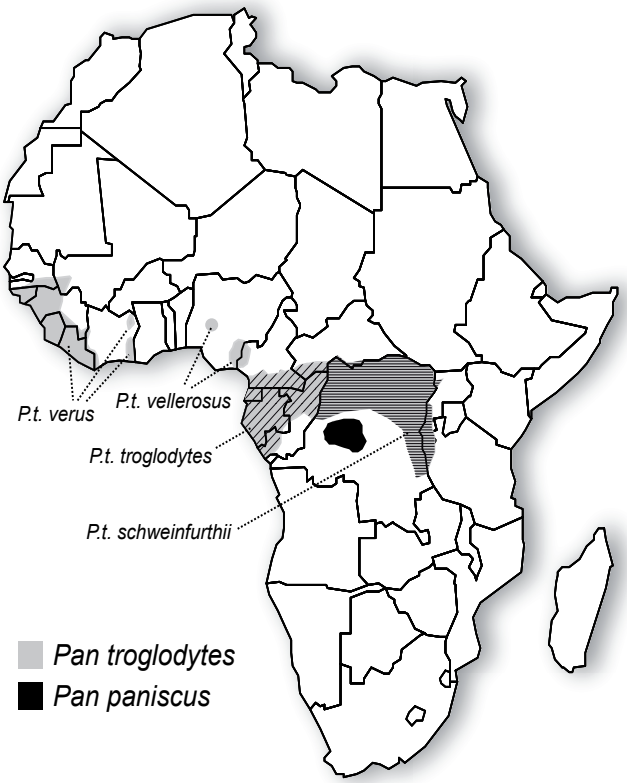


Fig. 2. Habitats of the different chimpanzee subspecies (*Pan troglodytes*) and the pygmy chimpanzee/bonobo (*Pan paniscus*).

normal tissues, and that these antigens differ among individuals of the same species or of different inbred strains. An exchange of tissue between such individuals would provoke an immune response, resulting in the rejection of the transplant. In 1948, Gorer together with his colleague Snell proposed naming the tumor resistance factors "histocompatibility (H) genes", with individual genes being distinguished by serial numbers. The first identified histocompatibility gene was named H-2, which was derived from its original name "antigen II" (8). The discovery of the human MHC can be attributed to J. Dausset, R. Payne and J. van Rood. Dausset detected leukocyte-agglutinating antibodies in sera of multitransfused patients (9), whereas Payne and van Rood observed the same phenomenon in multiparous women (10, 11). It was shown that human leukocyte antigens were controlled by one major complex of genes, designated HLA, which appears to be homologous to the mouse H-2 system. To date, the major histocompatibility complex has been described for numerous vertebrate species.

The human MHC region

The MHC is defined as a group of genes encoding for cell-surface proteins primarily responsible as the context elements for the recognition of degraded foreign antigens by T lymphocytes. In most vertebrate species, it is characterized by a large number of genes, some of which display an extensive degree of polymorphism (12, 13). The human MHC has been mapped on the short arm of chromosome 6 (14), and can be divided into a class I, II, and III region (Fig. 3).

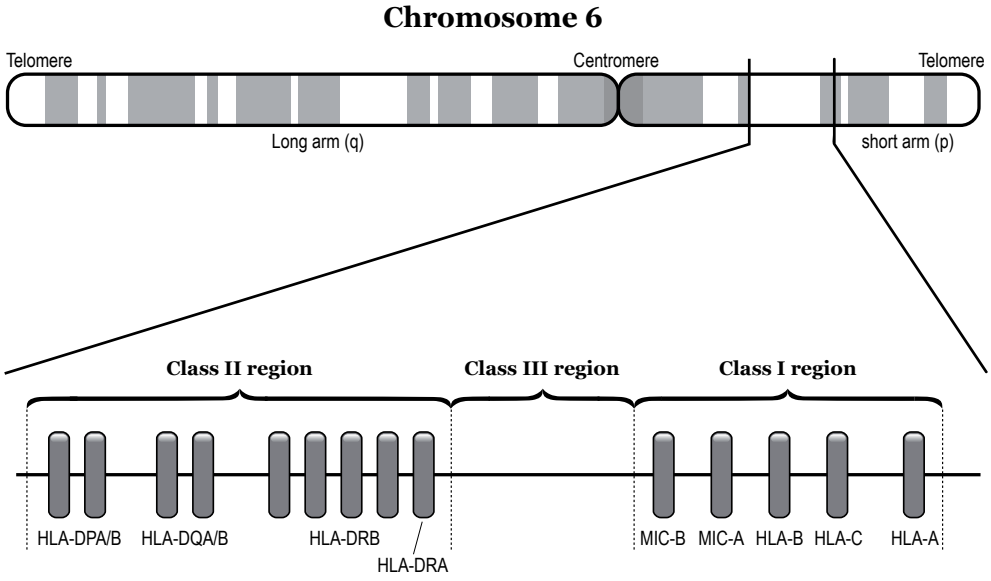


Fig. 3. Genomic map of the human MHC region.

The *Mhc* class I region is located telomeric on the short are of chromosome 6. In humans two groups of MHC class I molecules are distinguished: classical and non-classical. The classical molecules are designated HLA-A, -B, and -C (15), and are composed of a heavy chain that complexes with β_2-microglobulin: they are expressed on virtually all nucleated cells, and present peptides of intracellular origin of approximately 9 amino acids in length (16). The complex of an MHC class I molecule together with a foreign antigen can be recognized by cytolytic $CD8^+$ T cells that can then lysis the cell. The non-classical molecules, HLA-E, -F, and -G, are known to be less polymorphic, and may show a restricted tissue distribution and a specialized function. HLA-E, and -G play a role in the inhibition of natural killer (NK) cell-induced lysis (17-19), whereas the function of HLA-F is unclear.

However, HLA-F tetramers were shown to bind immunoglobulin-like transcript-2 (ILT2) and ILT4 receptors (20). Although in the cells studied to date HLA-F appears to remain within the cell, the tetramer study suggests that in particular circumstances this gene product may reach the cell surface.

The *Mhc* class II region is located more centromeric on the short arm of chromosome 6, and encodes three main isotypes: HLA-DP, -DQ, and -DR (Fig. 3). The molecules are heterodimers composed of an alpha and a beta chain encoded by the *A* and *B* genes, respectively. In contrast to MHC class I molecules, they usually bind peptides of an extracellular origin, 13-25 amino acids in length (21). The MHC class II molecules are expressed primarily on white blood cells, and control antibody responses and $CD4^+$ T-cell-mediated help. In addition, HLA-DM and -DO, two MHC class II-like molecules, are involved in peptide loading and transport of MHC class II molecules (22).

The MHC class III region is situated between class I and II. It contains a variety of genes coding for molecules with different immunological functions: for instance, genes that are involved in the innate immunity, or in inflammation and the regulation of immunity. The MHC class III region measures approximately 700 kb, and is the most gene-dense region known on the human genome (23).

The MHC of chimpanzees: A comparison with HLA

The MHC of chimpanzees was discovered in 1974 (24), when it was designated as ChLA, and then later renamed *MhcPatr* (25). As in humans, the chimpanzee MHC region is divided into a class I, II, and III region. In the early days, chimpanzee MHC polymorphism was defined based on serological assays, but at present a tremendous amount of molecular data is on hand. A comparison of the complete *Mhc* class I region between humans and chimpanzees is available (26). The genomic organization of a particular *Patr-DR* region configuration was mapped by using two overlapping cosmids (27).

The classical class I loci in chimpanzees are *Patr-A*, *-B*, and *-C* (25). They are composed of eight exons, and most of the polymorphism is confined to exon 2 and 3, coding for the peptide-binding site of the molecule (28). Phylogenetic analysis showed that alleles (variant forms of a gene) with a shared ancestry cluster into lineages. Some of these lineages may even predate the speciation of human and chimpanzees, the so-called trans-species lineages (29, 30). *HLA-A* locus alleles can be divided into two lineages comprising six families. The separation is based on the substitution patterns of 33 diagnostic nucleotide positions. The HLA-A2 lineage contains the *HLA-A2*, *-A10*, and *-A19* families, whereas the HLA-A3 lineage includes the *HLA-A9*, *-A1/A3/A11/A30*, and *-A80* families (31, 32). Notably, the relatively small sample of chimpanzees analyzed only possesses *Patr-A* alleles that cluster into the *HLA-A1/A3/A11/A30* family, and members of the other five families appear to be absent (30). The variation seen in the *A* locus alleles is mostly due to point mutations. The variation in the *HLA-B* locus is caused by a frequent exchange of polymorphic sequence motifs between lineages (33-35), and this phenomenon is also observed for the *Patr-B* locus (34). The *HLA-C* locus can be divided into three lineages, *HLA-C3*, *-C7*, and *-C17* (36).

The chimpanzee has representatives in the *HLA-C3* and *-C7* lineages, and, in addition, has a species-specific lineage, *Patr-C2*. The *HLA-C17* lineage is human specific. In general, the variation observed for the *C* locus is, like for the *A* locus, mainly based on point mutations (37). The MHC class II molecules in chimpanzees are designated Patr-DP, -DQ, and -DR (25). A func-tional *Mhc* class II gene is composed of six exons. Most polymorphism is confined to exon 2, which codes for an important part of the peptide-binding site of the molecule (38). Humans and chimpanzees show a similar organization of the *DP* region containing four genes, *DPA1*, *DPA2*, *DPB1*, and *DPB2*. Only the *DPA1*/*DPB1* pair is expressed, while *DPA2* and *DPB2* are pseudogenes, possessing multiple deleterious mutations (39). The *Patr-DQ* region also possesses a duplicated set of genes (40). Based on an analysis of the promoter region, *Alu* repeats, and endogenous retroviruses it is postulated that humans and chimpanzees have a similar organization of the *DQ* region (41, 42). Both species share particular *DQA1* and *DQB1* lineages, but also possess species-unique lineages (43). For example, the trans-species *Mhc-DQB1*15* lineage, which is present in chimpanzees and rhesus macaques, and probable lost during human evolution (41). The *HLA-DQA2*/*DQB2* gene pair is not expressed. A comparison of alleles at these loci revealed that they are highly similar between humans and chimpanzees (41, 42, 44, 45). The *DR* region is, as compared to the *DP* and *DQ* regions, far more complex. Several *HLA-DRB* loci are characterized, and were probably formed by independent duplications (46). The equivalents of the *HLA-DRB1-DRB9* genes and gene segments are detected in chimpanzees (41). Like in humans, the *Patr-DRB2*, and *-DRB6-DRB9* loci are pseudogenes. These loci lack complete exons, contain frame shift deletions, and/or have premature stop codons (47-51). Although the *DRB6* locus lacks exon 1 and is considered to be a pseudogene, it is transcribed at moderate levels by using the promoter of a retroviral insert (52). Chimpanzees possess two additional loci, *Patr-DRB*W8* and *-DRB*W9*[1], which are also present in gorillas but absent in humans. Phylogenetic analysis indicated that these loci are closely related to the *DRB3* locus (53, 54). The *HLA-DRA* locus is oligomorphic, and two alleles are described (55, 56), a situation that is mirrored by the *Patr-DRA* locus (57, 58). The number of *DRB* genes present per haplotype/region configuration[2] differs, and can range from three to five in humans, and from three to six in chimpanzees. Five and eight different region configurations are described in humans and chimpanzees (27, 50, 59). Only the organization of the *HLA-DR7* region configuration is shared between the two species (48, 50, 60). The chimpanzee haplotypes were described by contig mapping or deduced from segregation studies: the order of most human *DRB* genes was determined by genomic mapping.

[1]The *W* stands for Workshop. The nomenclature is used when it is uncertain which locus controls the alleles (25).

[2]Haplotype/region configuration: unique combination of MHC genes/alleles present on a chromosome.

MIC genes in humans and chimpanzees

MIC stands for major histocompatibility complex class I chain-related gene, and is located on chromosome 6, within the MHC class I region (Fig. 3) (61, 62). In humans, seven *MIC* genes are distinguished; *MICA* and *B* are functional genes, while *MICC* to *G* are pseudogenes. The *MICA* and *B* genes show a high degree of similarity to the classical *Mhc* class I genes, but differ in their unusual organization of exons-introns. The molecules do not associate with β_2-microglobulin, interferon γ does not up-regulate the expression, and they seem to be dominantly expressed on fibroblast and epithelial cells (63). Cell stress is an important stimulator for the up-regulation of the molecules on the cell surface, and they react preferentially with the ligands Vδ1 $\gamma\delta$ TCR and NKG2D to induce an immune reaction (64-66). Moreover, both genes are found to display polymorphism in humans. In chimpanzees, thus far, only one functional *MIC* gene has been described (67), which has an intermediate character as compared to the human *MICA* and *B* genes.

Microsatellites

Microsatellites are defined as tandem repeats of short sequence motifs, which are 2 to 6 base pairs long. They are evenly distributed throughout the genome and are highly polymorphic (68, 69). Microsatellites are present in all organisms that have been investigated to date (70), although they are more common in eukaryotes than in prokaryotes (71). Varying the number of repeat units generates polymorphism/allelic variation, and generally up to 30 repeats are detected (70). Compared to the rate of point mutations (10^{-9} to 10^{-10} per site per year), the mutation rate of microsatellites is substantially higher. In humans, this rate is estimated to be around 10^{-3} events per microsatellite locus per generation (72, 73). The high mutation rate in microsatellites can be explained in two ways: first, mutations can occur through replication slippage (slipped strand mispairing), which results in the gain or loss of repeat units (70, 74, 75); second, recombination can play a role in changing the lengths of microsatellites via unequal crossing-over or gene conversion[1] (70).

The poly(A)/poly(T) is the most common mononucleotide repeat in the human genome, but unsuitable for mapping or for population analysis due to its instability during PCR amplification (76-78). The most frequent dinucleotide repeat is the CA/TG repeat (76), whereas CAG and AAT are the most common trinucleotide repeats (79, 80). At present, microsatellites are often used as genetic linkage markers in genome mapping projects (81), and for evolutionary and population studies in many different species (82, 83).

[1]Unequal crossing-over: recombination event occurring between two sites that are not homologous, resulting in a deletion in one DNA strand and an insertion in the other strand. Gene conversion: nonreciprocal transfer of information from one chromatid (which remains unchanged) to another.

Human immunodeficiency virus type 1 (HIV-1)

HIV-1 features and origin

The human immunodeficiency virus type 1 (HIV-1) was first identified in 1983 (84), and is the virus that initiated the death of many members of gay communities in the USA two years earlier. The disease, now named acquired immunodeficiency syndrome (AIDS), is characterized by a decline in $CD4^+$ T cells, opportunistic infections (for instance, *Pneumocytis carinii* pneumonia), neurological abnormalities, gastrointestinal disorders, and malignancies (Kaposi's sarcoma) due to an affected immune system (85, 86). HIV-1, and with that AIDS, originated on the African continent. However, only years after the discovery of the virus it was recognized in Africa. Later, the virus spread to Europe and Asia as well. By the end of 2008, 33.4 million people worldwide were infected with HIV-1: about 2.5 million people were newly infected during that year, and 2 million people died in that year as a result of the disease (UNAIDS, 2009).

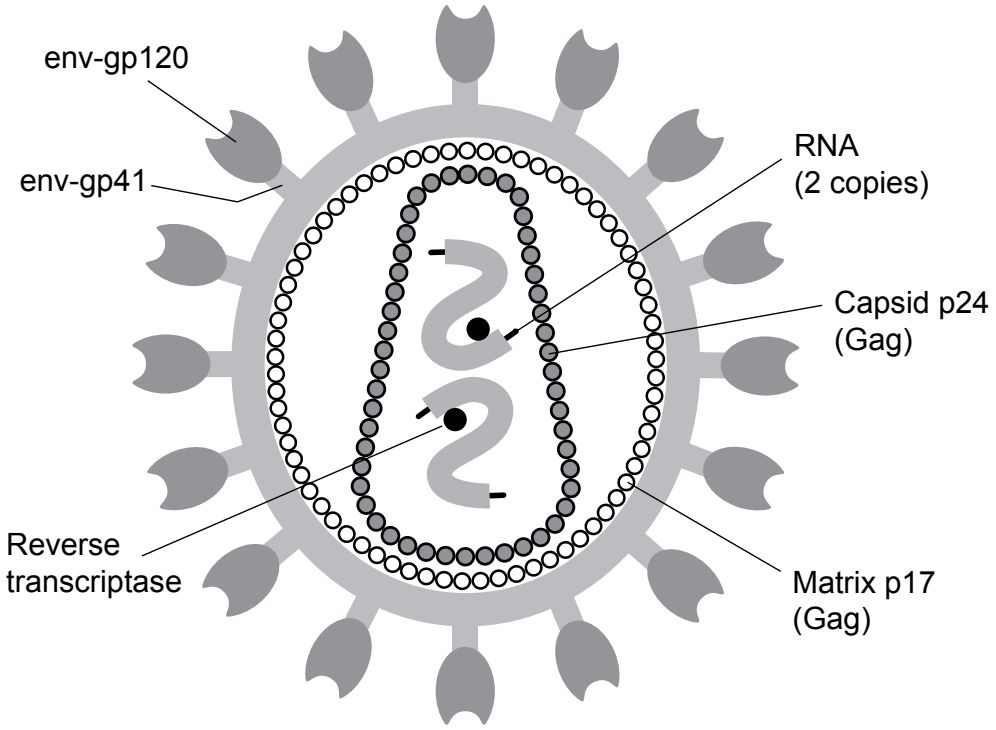


Fig. 4. Structure and composition of an HIV-1 particle. Env-gp120 is the surface glycoprotein, and Env-gp41 is the transmembrane protein.

HIV-1 belongs to the family of retroviruses (*Retroviridiae*) and to the group of lentiviruses (*Lentiviridiae*). Retroviruses are enveloped, positive single-stranded RNA viruses, and each virus particle contains two copies of the virus genome (Fig. 4). They are the smallest virus species known, with a genomic length varying from 3.5 to 9 kb, and have the unique capability to synthesize DNA from an RNA template. Lentiviruses usually have a long incubation time, sometimes causing a chronic and lifelong infection in their host. Their name is derived from this property, as the Latin word lentus means slow. They are known to cause immunodeficiency, disorders in the haematopoietic and central nervous system, and sometimes arthritis or autoimmunity (87). Different species seem to harbor lentiviruses: for instance, horses (equine infectious anemia virus, EIAV), cattle (bovine immunodeficiency virus, BIV), cats (feline immunodeficiency virus, FIV), goats (caprine arthritis-encephalitis virus, CAEV), sheep (maedi-visna virus, MVV), and primates (human or simian immunodeficiency virus, HIV/SIV). Based on nucleotide sequences, HIV-1 isolates are classified into three different groups: named M, N, and O. Group M comprises the "main" HIV-1 isolates, and is divided into eleven subtypes (A-K). Group O contains the "outliers", and group N is based on a virus isolate from Cameroon (YBF30). This latter virus seems to be more related to the M than to the O group (Fig. 5) (88, 89). Recently a new HIV-1 lineage, closely related to the gorilla simian immunodeficiency virus (SIV_{gor}), has been distinguished in a Cameroonian individual (isolate RBF168), and was proposed to be designated HIV-1 group P (Fig. 5) (90).

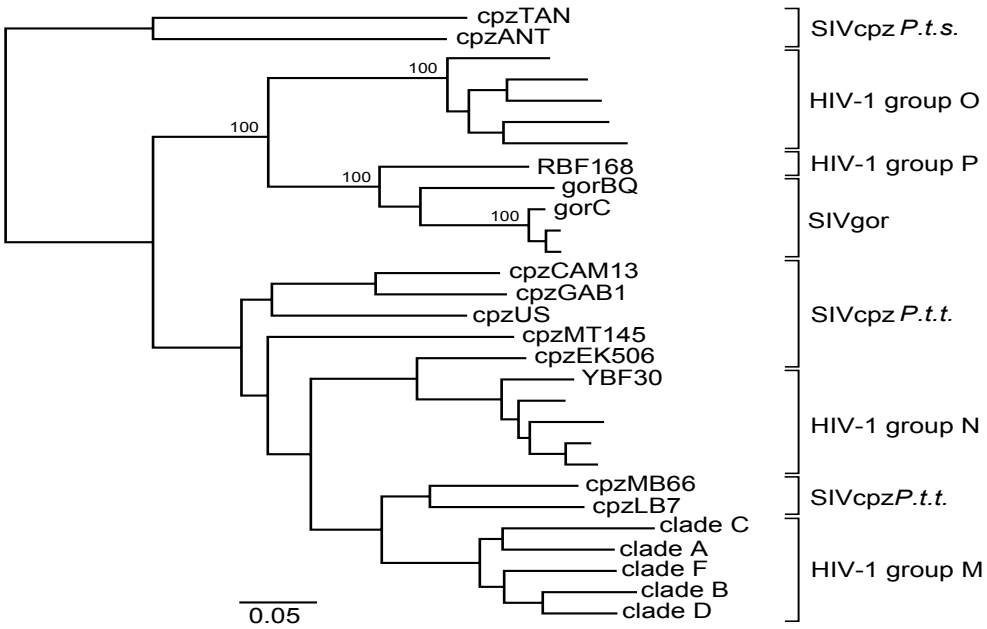


Fig. 5. Phylogenetic comparison of different HIV-1 and SIV_{cpz}/SIV_{gor} strains. Brackets indicate the groups in which the different HIV-1/SIV sequences can be classified. *P.t.s.*: *Pan troglodytus schweinfurthii*. *P.t.t.*: *Pan troglodytes troglodytus*. (Adapted from Plantier J.C. et al., Nature Medicine 15, 871-872, 2009).

The source of the HIV-1 infection in humans is the chimpanzee simian immunodeficiency virus (SIV_{cpz}). SIV_{cpz} was first characterized in 1990 (91). Genetic analysis on four different SIV_{cpz} strains showed that the HIV-1 groups M, N, and O originated from at least three independent introductions of SIV_{cpz} into the human population (88). These zoonotic events probably occurred as a result of bushmeat hunting (88, 89). At least one of the SIV_{cpz} strains arose from cross-species transmission and recombination events from SIV-infected monkey species (red-capped mangabeys (rcm), *Cercocebus torquatus*, and greater spot-nosed monkeys (gsn), *Cercopithecus nictitans*) on which chimpanzees prey. This is illustrated by phylogenetic analysis, which shows that the Pol sequence of SIV_{cpz} clusters tightly to SIV_{rcm}, whereas the Env sequence of SIV_{cpz} clusters closely to SIV_{gsn} (Fig 6) (92, 93). The origin of the pandemic with HIV-1 group M is thought to be a viral lineage that persists today in *P.t.troglodytes* in southeastern Cameroon. HIV-1 group N is probably derived from a second SIV_{cpz} lineage in *P.t.t.* in south central Cameroon, and seems to be more restricted to this area (94). The origin of HIV-1 group O remains uncertain. Group O-like viruses are documented in wild gorillas (SIV_{gor}) (95, 96). There is indication that *P.t.t.* animals are the source of SIV_{gor} (97). However, gorillas are known to be herbivores, and since physical encounters with chimpanzees are rare, the question remains as to how gorillas acquired their SIV, and whether gorillas and/or chimpanzees are responsible for the HIV-1 group O zoonosis (96, 97).

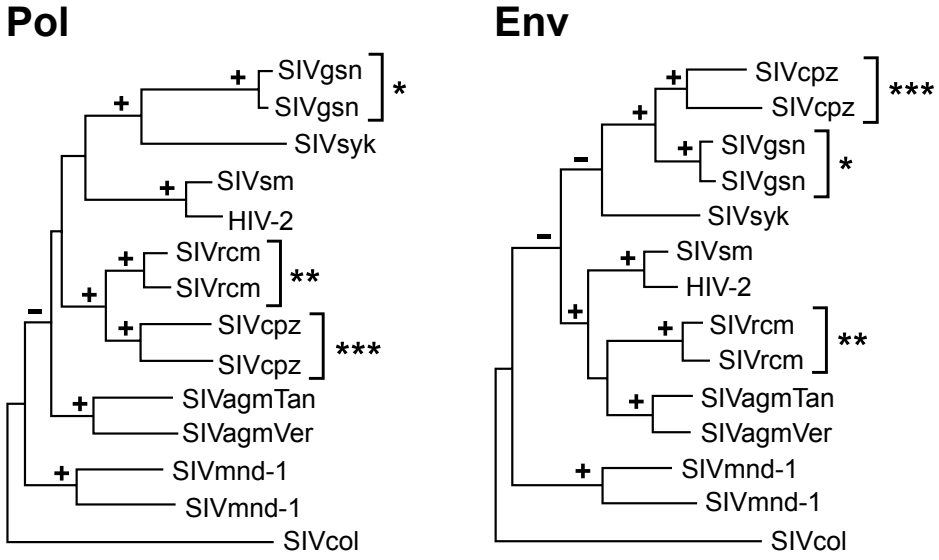


Fig. 6. Phylogenetic comparison of the Pol and Env sequences of different SIV strains. The greater spot-nosed monkey (*), red-capped mangabeys (**), and chimpanzee (***) derived simian immunodeficiency viruses are indicated. A - and + indicate internal branches found in at least 70% and 95% of the bootstrap replicates (adapted from Bailes E. et al., Science 300, 1713, 2003).

Along with humans, chimpanzees can be infected with HIV-1. In the past, over 150 chimpanzees worldwide have been experimentally infected with HIV-1, but only one animal developed symptoms of AIDS (98). Hence, in general HIV-1 infected chimpanzees are relatively resistant to progression to AIDS, which makes this a worthwhile phenomenon to investigate.

Genomic structure of HIV-1/SIV$_{cpz}$

Compared to other retroviruses, lentiviruses have a larger genome containing considerable more genes. All retroviruses contain Gag, Pol, and Env, three functional proteins that are essential for the construction of the virus particle (Fig. 7). The Gag protein codes for the intra-viral proteins; Pol codes for the enzymes reverse transcriptase (synthesizes DNA from RNA), integrase (necessary for the integration of virus DNA into the host genome), and protease (cutting the Gag-Pol polyprotein into functional subunits); and Env codes for the membrane proteins (gp120 for the surface membrane protein, and gp41 for the transmembrane protein). Both the 5' and 3' side of the genome possess a long terminal repeat (LTR). The LTRs contain sequences that are important for the initiation of the transcription and termination, reverse transcriptase, integrase, and binding of the transactivator Tat. In addition to this common structure of retroviruses, lentiviral genomes encode a number of auxiliary (Tat and Rev) and accessory genes (Vif, Vpr, Vpu, and Nef) (Fig. 7). The Tat and Rev proteins have a regulatory function, and their presence is essential for the replication of the virus. The accessory proteins are present on the viral messenger RNA (mRNA), but they are not required for the replication of the virus *in vitro*. However, *in vivo* these proteins are necessary for the replication and virulence of the virus (99, 100).

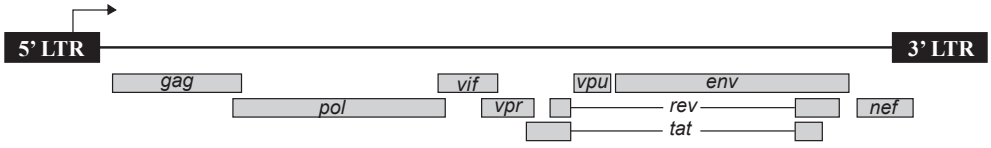


Fig. 7. Genomic organization of HIV-1/SIV$_{cpz}$. The arrow indicates the transcription initiation site (adapted from Singh, Virology Journal 3, 60, 2006).

HIV-1 evasion and escape from the immune system

Humans possess an innate and adaptive immune system to protect themselves from infections. However, the HIV/SIV virus employs different "innate" and "adaptive" strategies to evade the immune system's various defenses.

The different innate strategies used by the virus can be categorized into a group that evades the antibody responses, and one that evades the cellular responses (101). Envelope glygoproteins of HIV/SIV have evolved structural properties that make them inherently resistant

to neutralizing antibodies. These are based on biochemical properties of glycosylation and oligomerisation of the virus envelope, and on the induction of conformational changes[1]. The number of glycosylation sites varies between different HIV isolates: about 24 N-linked and eight O-linked[2] glycosylation sites are observed in gp120, and four N-linked sites in gp41 (102, 103). The sites are found especially on the surface of the outer domain of gp120 (104), and therefore this part is less accessible for neutralizing antibodies. Nevertheless, neutralizing antibodies are found against primary HIV-1 isolates, which are directed against conserved parts of the CD4 binding site and the membrane section of gp41. The above-mentioned observations are currently used in vaccine strategies to increase the immunogenicity of the envelope glycoprotein complexes (101).

A strategy to evade the cellular immune response is to diminish the T-helper function. A subclinical loss of $CD4^+$ T cells already results in reduced T-helper function, which leads to consequences in different directions of the immune system (Fig. 8). HIV-1 also has the capability to kill infected $CD4^+$ T lymphocytes with a killing of non-infected bystander $CD4^+$ and $CD8^+$ T cells via receptor-programmed cell death. Receptor-ligand combinations that are involved are Fas/Fas-L and TNFα/TNFRII. For instance, the accessory gene product Nef stimulates the up-regulation of Fas-L on HIV-1-infected $CD4^+$ T cells and macrophages. Fas-L interacts with Fas expressed on non-infected $CD4^+$ and $CD8^+$ T cells, resulting in the apoptosis of these cells (Fig. 8) (101). Furthermore, HIV-1 can circumvent the cellular immune system by persistent and latent-like infections of inert long-lived cells, because the immune system is incapable of eliminating resting cells. Moreover, the infection of cells in the central nervous system (CNS) is another circumvention pathway, as generally T cells have limited access to the CNS. In addition, HIV-1 can evade by the use of different co-receptors to enter $CD4^+$ T cells. CC-chemokine receptor-5 (CCR5) is used by nonsyncytium-inducing (NSI), macrophage-tropic viruses, whereas C-X-C-chemokine receptor-4 (CXCR4) functions as a co-receptor for T cell line-adapted, syncytium-inducing (SI) virus strains. In vitro studies demonstrated that primary HIV-1 isolates primarily use CCR5 as co-receptor to enter the cell. Later in the disease progression, the virus uses different co-receptors. This change in coreceptor usage is associated with a switch from NSI to SI strains, a loss of sensitivity for chemokines, and a decrease in the number of $CD4^+$ T cells (105). Finally, the ability of HIV-1 to down-regulate the expression of MHC class I molecules by the accessory gene product Nef is an alternative feature to escape the cellular immune response (Fig. 8). This down regulation makes the infected cells less detectable for cytotoxic T cells (CTL). Nef in particular down-regulates the MHC class I A and B molecules, but does not affect the NK cell-activating MHC class I C and E molecules (106).

[1]Glycosylation: modification of a molecule by adding sugars. Oligomerisation: folding of several proteins to form a molecule. Conformational changes: interactions between molecules that results in a change of shape of one of the molecules.

[2]N-linked glycosylation: enzymatic process that adds oligosaccharides to asparagine; O-linked glycosylation: enzymatic process that adds oligosaccharides to serine or threonine.

HIV-infected individuals have increased levels of NK cells expressing HLA-C-specific inhibitory NK cell receptors (iNKRs), which prevent the killing of infected $CD4^+$ T cells (107, 108). However, despite all these cellular immune evasion strategies of HIV/SIV, there are infected individuals that are capable of producing antibodies and CTL responses that temporarily diminish the virus replication and delay the onset of the disease (109).

The "adaptive" approach of HIV/SIV to circumvent the immune system is via the selection of immune escape variants, which, in addition to the innate strategies, plays an important role in AIDS pathogenesis. The genetic diversity of HIV-1 is generated by a high replication rate of the virus (approximately 10^9 new virus particles per day) in combination with a disposition of the enzyme reverse transcriptase to create errors (mutation frequency is $3x10^{-5}$ per nucleotide base per cycle of replication) (110). Mutations are created within the envelope glycoproteins of HIV-1 to escape the virus-specific antibody responses (111, 112). This is accomplished via changes in the amino acid sequence of the epitopes that are recognized, or through conformational changes due to amino acid substitutions elsewhere in the molecule (113-115). Mutations within or nearby the CTL epitope occur as well, and allow HIV-1 to escape CTL responses (116-118). This type of escape often occurs through substitutions within key anchor residues, which results in the absence or reduced binding of the peptide to the MHC class I molecule. This process is usually irreversible due to a loss of presentation of the viral peptide (119, 120). Changes in residues that contribute to T-cell receptor recognition are also detected (118). However, the mutated epitopes are ultimately recognized by subsequent CTL responses, and escape is not permanent. Moreover, the type of T-cell receptor that is expressed on the HIV-1-specific CTL is important, and indicates in which phase of the infection CTL escape will take place (121). For other type of viruses CTL escape variants exist that interfere with the intracellular processing. Amino acid substitutions that occur in the flanking residues of the epitopes are probably important for proteosome cleavage, and contribute to the escape. This phenomenon has recently been described for HIV-1 as well (122, 123), and appears more frequent than was suspected (124). Escape variants in the acute as well as in the chronic phase of an HIV-1 infection are observed (121). The rate at which they occur relies on the selective pressure on a particular epitope, and on the amount of sequence variation a virus tolerates within the epitope-coding region. The selective pressure depends on the strength of the immune response, the virus antigen, and the avidity of TCR recognition (101).

Thesis aim and outline

Chimpanzees and bonobos seem to possess *Mhc* class I *A* locus alleles that only cluster into the *HLA-A1/A3/A11/A30* family, and members of the five other families present in humans are absent in these two species (30). However, a thorough population study is lacking. In this thesis, the *Mhc* class I repertoire of the pedigreed outbred chimpanzee colony housed at the BPRC was investigated (**Chapter 2**).

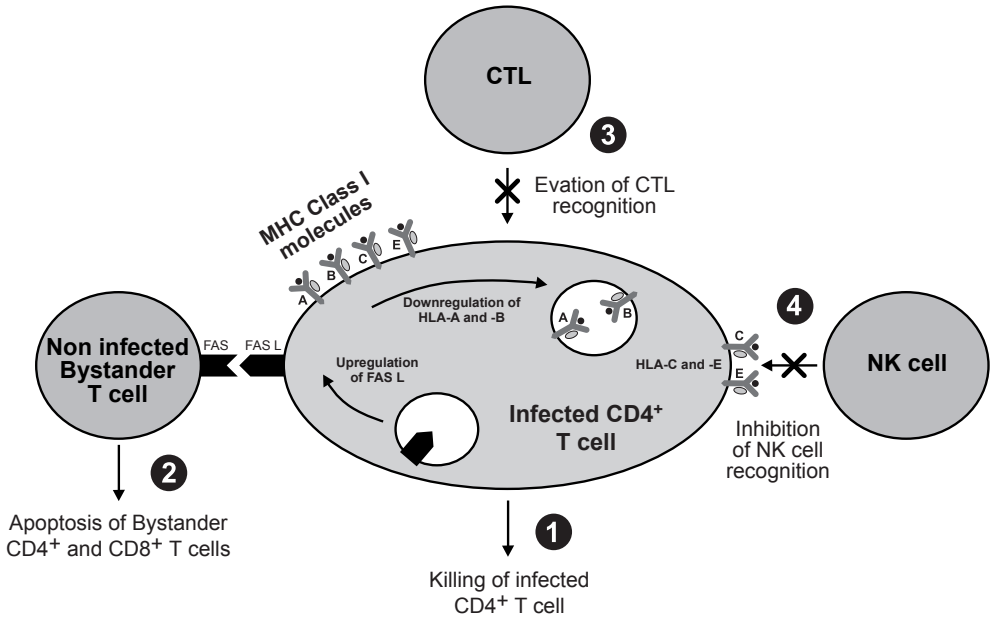


Fig. 8. Overview of the different innate strategies that HIV-1 uses to evade cellular immune responses. (1) Killing of infected $CD4^+$ T cells results in the disruption of T helper-cell responses that are needed to induce virus-specific antibodies and CTL responses. (2) Receptor-programmed cell death results in the killing of non-infected bystander $CD4^+$ and $CD8^+$ T cells, and thereby interfers in the development of virus-specific T helper-cell and CTL responses. (3) Nef down-regulation of HLA-A and -B molecules interferes in the recognition and killing of infected cells by virus-specific CTL. (4) Nef regulates the continuous expression of HLA-C and -E molecules on the virus-infected cell, resulting in the inhibition of NK cell recognition (adapted from D.T. Evans and R.C. Desrosiers, Immunological reviews 183, 141-158, 2001).

Mhc class I and II sequences encode for molecules that are under frequency-dependent/diversifying selection, which makes it difficult to make firm conclusions with regard to the loss of alleles/lineages, for instance, due to disease susceptibility. To investigate whether negative/purifying selection acted on the *Patr* class I loci, intron 2 of the different class I alleles, located between the two most polymorphic exons, was sequenced (**Chapter 3**). The results evidenced that the *Mhc* class I repertoire of chimpanzees was reduced by a selective sweep, and it was put forward that the cause may have been HIV-1/SIV_{cpz} or a related ancestral retroviral infection. This hypothesis forms the basis of the additional reports in this thesis.

We determined whether the *Mhc* class I repertoire reduction extends to other regions on chromosome 6. Therefore, a study investigating genetic variation involving the major histocompatibility complex class I chain-related gene *A* (*MICA*), known to be polymorphic

in humans and located next to the *Mhc* class I *B* locus, was performed (**Chapter 4**). In addition, a different approach was undertaken by performing a multi-locus comparison between humans and chimpanzees using microsatellite markers located in the MHC region and involving a variety of other chromosomes (**Chapter 5**). Chimpanzees are known to display more allelic variation in their mtDNA (6, 125), and in particular nuclear genes than do humans (126-128). This is consistent with the knowledge that chimpanzees as a species are older than modern humans and have existed as more subdivided populations, which resulted in more genetic variation. If the chimpanzee MHC class I region had undergone selection, limited variation involving the MHC microsatellite markers as compared to humans would have been expected. In contrast, however, the markers located on other chromosomes are expected to show more or at least equal amounts of variation in chimpanzees with regard to humans. Subsequently, haplotype diversity in the chimpanzee population was studied (**Chapter 5**).

A recent study in humans indicated that DRB-specific $CD4^+$ T-cell responses are playing an important role in resistance/susceptibility to HIV-1 infection (129). This observation initiated the detailed characterization of the *Patr-DRB* region. It was investigated which of the alleles present on a haplotype are transcribed, and may play a role in providing immunological help to $CD8^+$ memory T cells (**Chapter 6**).

Particular MHC class I molecules in HIV-1-infected humans are associated with long-term non-progression. These molecules, specified as HLA-B*27 and -B*57, present both conserved Gag-specific peptides to CTL (130-132). The *Mhc* class I repertoire of chimpanzees experienced a selective sweep, and it was examined as to whether it reflects reality that HIV-1/SIV_{cpz} or a related ancestral retrovirus was the cause. For different *Patr* class I molecules was investigated whether they have the capacity to bind conserved HIV-1/SIV_{cpz} Gag peptides that probably contribute to control viral replication, and thus help in the prevention of progression to AIDS in chimpanzees (**Chapter 7**).

Finally, the data are summarized, and the observed repertoire reduction in relation to the relative resistance of chimpanzees to developing AIDS is put into a wider context (**Chapter 8**).

References

1. Groves CP (2001) Primate Taxonomy (Smithsonian Institution).
2. Martin RD (1990) Primate origins and evolution (Chapman and Hall London).
3. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
4. Fujiyama A, Watanabe H, Toyoda A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
5. Sibley CG, Ahlquist JE (1987) DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J Mol Evol 26:99-121.
6. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
7. Gonder MK, Oates JF, Disotell TR, et al. (1997) A new west African chimpanzee subspecies? Nature 388:337.
8. Klein J (1986) Natural History of the Major Histocompatibility Complex (John Wiley and Sons).
9. Dausset J (1954) leuco-agglutinins IV; leoco-agglutinins and blood transfusion. Vox Sanguinis 4:190-198.
10. Payne R, Rolfs MR (1958) Fetomaternal leukocyte incompatibility. J Clin Invest 37:1756-1763.
11. Van Rood JJ, Eernisse JG, Van Leeuwen A (1958) Leucocyte antibodies in sera from pregnant women. Nature 181:1735-1736.
12. Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67-74.
13. Riley E, Olerup O (1992) HLA polymorphisms and evolution. Immunol Today 13:333-335.
14. Francke U, Pellegrino MA (1977) Assignment of the major histocompatibility complex to a region of the short arm of human chromosome 6. Proc Natl Acad Sci U S A 74:1147-1151.
15. Bodmer JG, Marsh SG, Parham P, et al. (1990) Nomenclature for factors of the HLA system, 1989. Tissue Antigens 35:1-8.
16. Falk K, Rotzschke O, Stevanovic S, et al. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290-296.
17. Braud VM, Allan DS, O'Callaghan CA, et al. (1998) HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. Nature 391:795-799.
18. Perez-Villar JJ, Melero I, Navarro F, et al. (1997) The CD94/NKG2-A inhibitory receptor complex is involved in natural killer cell-mediated recognition of cells expressing HLA-G1. J Immunol 158:5736-5743.
19. Soderstrom K, Corliss B, Lanier LL, Phillips JH (1997) CD94/NKG2 is the predominant inhibitory receptor involved in recognition of HLA-G by decidual and peripheral blood NK cells. J Immunol 159:1072-1075.
20. Allan DS, Lepin EJ, Braud VM, et al. (2002) Tetrameric complexes of HLA-E, HLA-F, and HLA-G. J Immunol Methods 268:43-50.
21. Brown JH, Jardetzky T, Saper MA, et al. (1988) A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature 332:845-850.
22. van Ham SM, Tjin EP, Lillemeier BF, et al. (1997) HLA-DO is a negative modulator of HLA-DM-mediated MHC class II peptide loading. Curr Biol 7:950-957.
23. Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122.

24. Balner H, Gabb BW, D'Amaro J, et al. (1974) Evidence for two linked loci controlling the serologically defined leukocyte antigens of chimpanzees (ChL-A). Tissue Antigens 4:313-328.
25. Klein J, Bontrop RE, Dawkins RL, et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217-219.
26. Anzai T, Shiina T, Kimura N, et al. (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc Natl Acad Sci U S A 100:7708-7713.
27. Brandle U, Ono H, Vincek V, et al. (1992) Trans-species evolution of *Mhc-DRB* haplotype polymorphism in primates: organization of *DRB* genes in the chimpanzee. Immunogenetics 36:39-48.
28. Bjorkman PJ, Saper MA, Samraoui B, et al. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512-518.
29. Lawlor DA, Ward FE, Ennis PD, et al. (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268-271.
30. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
31. Domena JD, Hildebrand WH, Bias WB, Parham P (1993) A sixth family of *HLA-A* alleles defined by *HLA-A*8001*. Tissue Antigens 42:156-159.
32. Kato K, Trapani JA, Allopenna J, et al. (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 143:3371-3378.
33. Belich MP, Madrigal JA, Hildebrand WH, et al. (1992) Unusual *HLA-B* alleles in two tribes of Brazilian Indians. Nature 357:326-329.
34. McAdam SN, Boyson JE, Liu X, et al. (1994) A uniquely high level of recombination at the *HLA-B* locus. Proc Natl Acad Sci U S A 91:5893-5897.
35. Watkins DI, McAdam SN, Liu X, et al. (1992) New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature 357: 329-333.
36. Wells RS, Seielstad MT, Bunce M, et al. (1997) *Cw*1701* defines a divergent african *HLA-C* allelic lineage. Immunogenetics 46:173-180.
37. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51:410-424.
38. Marsh SG, Bodmer JG (1989) HLA-DR and -DQ epitopes and monoclonal antibody specificity. Immunol Today 10:305-312.
39. Trowsdale J, Kelly A, Lee J, et al. (1984) Linkage map of two HLA-SB beta and two HLA-SB alpha-related genes: an intron in one of the SB beta genes contains a processed pseudogene. Cell 38:241-249.
40. Bontrop RE, Otting N, Broos LA, et al. (1989) RFLP analysis of the *HLA-*, *ChLA-*, and *RhLA-DQ* alpha chain gene regions: conservation of restriction sites during evolution. Immunogenetics 30:432-439.
41. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
42. Donner H, Tonjes RR, Bontrop RE, et al. (1999) Intronic sequence motifs of *HLA-DQB1* are shared between humans, apes and Old World monkeys, but a retroviral LTR element (DQLTR3) is human specific. Tissue Antigens 53:551-558.

43. de Groot NG, Bontrop RE (1999) The major histocompatibility complex class II region of the chimpanzee: towards a molecular map. Immunogenetics 50:160-167.
44. Gaur LK, Heise ER, Thurtle PS, Nepom GT (1992) Conservation of the *HLA-DQB2* locus in nonhuman primates. J Immunol 148:943-948.
45. Gaur LK, Hughes AL, Heise ER, Gutknecht J (1992) Maintenance of *DQB1* polymorphisms in primates. Mol Biol Evol 9:599-609.
46. Bodmer JG, Marsh SG, Albert ED, et al. (1995) Nomenclature for factors of the HLA system, 1995. Tissue Antigens 46:1-18.
47. Gongora R, Figueroa F, Klein J (1996) The *HLA-DRB9* gene and the origin of *HLA-DR* haplotypes. Hum Immunol 51:23-31.
48. Kenter M, Otting N, Anholts J, et al. (1992) *Mhc-DRB* diversity of the chimpanzee (Pan troglodytes). Immunogenetics 37:1-11.
49. Satta Y, Mayer WE, Klein J (1996) HLA-DRB intron 1 sequences: implications for the evolution of HLA-DRB genes and haplotypes. Hum Immunol 51:1-12.
50. Slierendregt BL, Kenter M, Otting N, et al. (1993) Major histocompatibility complex class II haplotypes in a breeding colony of chimpanzees (Pan troglodytes). Tissue Antigens 42:55-61.
51. Vincek V, Klein D, Figueroa F, et al. (1992) The evolutionary origin of the HLA-DR3 haplotype. Immunogenetics 35:263-271.
52. Mayer WE, O'HUigin C, Klein J (1993) Resolution of the HLA-DRB6 puzzle: a case of grafting a de novo-generated exon on an existing gene. Proc Natl Acad Sci U S A 90:10720-10724.
53. Bontrop RE, Otting N, Slierendregt BL, Lanchbury JS (1995) Evolution of major histocompatibility complex polymorphisms and T-cell receptor diversity in primates. Immunol Rev 143:33-62.
54. Slierendregt BL, Bontrop RE (1994) Current knowledge on the major histocompatibility complex class II region in non-human primates. Eur J Immunogenet 21:391-402.
55. Kralovicova J, Marsh SG, Waller MJ, et al. (2002) The *HLA-DRA*0102* allele: correct nucleotide sequence and associated HLA haplotypes. Tissue Antigens 60:266-267.
56. Lee JS, Trowsdale J, Travers PJ, et al. (1982) Sequence of an HLA-DR alpha-chain cDNA clone and intron-exon organization of the corresponding gene. Nature 299:750-752.
57. Bontrop RE, Elferink DG, Otting N, et al. (1990) Major histocompatibility complex class II-restricted antigen presentation across a species barrier: conservation of restriction determinants in evolution. J Exp Med 172:53-59.
58. Doxiadis GG, de Groot N, de Groot NG, et al. (2008) Reshuffling of ancient peptide binding motifs between *HLA-DRB* multigene family members: old wine served in new skins. Mol Immunol 45:2743-2751.
59. Bohme J, Andersson M, Andersson G, et al. (1985) HLA-DR beta genes vary in number between different DR specificities, whereas the number of DQ beta genes is constant. J Immunol 135:2149-2155.
60. Mayer WE, O'HUigin C, Zaleska-Rutczynska Z, Klein J (1992) Trans-species origin of Mhc-DRB polymorphism in the chimpanzee. Immunogenetics 37:12-23.
61. Bahram S, Bresnahan M, Geraghty DE, Spies T (1994) A second lineage of mammalian major histocompatibility complex class I genes. Proc Natl Acad Sci U S A 91:6259-6263.
62. Leelayuwat C, Townend DC, Degli-Esposti MA, et al. (1994) A new polymorphic and multicopy MHC gene family related to nonmammalian class I. Immunogenetics 40:339-351.
63. Bahram S (2000) MIC genes: from genetics to biology. Adv Immunol 76:1-60.
64. Bauer S, Groh V, Wu J, et al. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285:727-729.

65. Gleimer M, Parham P (2003) Stress management: MHC class I and class I-like molecules as reporters of cellular stress. Immunity 19:469-477.
66. Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279:1737-1740.
67. Steinle A, Groh V, Spies T (1998) Diversification, expression, and gamma delta T cell recognition of evolutionarily distant members of the MIC family of major histocompatibility complex class I-related molecules. Proc Natl Acad Sci U S A 95:12510-12515.
68. Amos B, Schlotterer C, Tautz D (1993) Social structure of pilot whales revealed by analytical DNA profiling. Science 260:670-672.
69. Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921.
70. Goldstein DBaS, C. (1999) Microsatellites: evolution and applications (Oxford University Press, Oxford).
71. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. J Mol Biol 293:151-160.
72. Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. Nat Genet 24:400-402.
73. Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2: 1123-1128.
74. Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-221.
75. Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211-215.
76. Beckman JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627-631.
77. Hancock JM (1996) Simple sequences and the expanding genome. Bioessays 18:421-425.
78. Stallings RL (1992) CpG suppression in vertebrate genomes does not account for the rarity of (CpG)n microsatellite repeats. Genomics 13:890-891.
79. Hancock JM (1995) The contribution of DNA slippage to eukaryotic nuclear 18S rRNA evolution. J Mol Evol 40:629-639.
80. Stallings RL (1994) Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. Genomics 21: 116-121.
81. Dib C, Faure S, Fizames C, et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152-154.
82. Bowcock AM, Ruiz-Linares A, Tomfohrde J, et al. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457.
83. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
84. Barre-Sinoussi F, Chermann JC, Rey F, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.
85. Drew WL, Mintz L, Miner RC, et al. (1981) Prevalence of cytomegalovirus infection in homosexual men. J Infect Dis 143:188-192.
86. Levy JA (1993) Pathogenesis of human immunodeficiency virus infection. Microbiol Rev 57:183-289.
87. Flint S, Enquist, LW, Krug, RM, Racaniello, VR, Skalka, AM (2000) Principles of Virology, molecular biology, pathogenesis, and control (ASM Press, Washington, DC).

88. Gao F, Bailes E, Robertson DL, et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature 397:436-441.
89. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287:607-614.
90. Plantier JC, Leoz M, Dickerson JE, et al. (2009) A new human immunodeficiency virus derived from gorillas. Nat Med 15:871-872.
91. Huet T, Cheynier R, Meyerhans A, et al. (1990) Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature 345:356-359.
92. Bailes E, Gao F, Bibollet-Ruche F, et al. (2003) Hybrid origin of SIV in chimpanzees. Science 300:1713.
93. Stanford CB, Wallis J, Matama H, Goodall J (1994) Patterns of predation by chimpanzees on red colobus monkeys in Gombe National Park, 1982-1991. Am J Phys Anthropol 94:213-228.
94. Keele BF, Van Heuverswyn F, Li Y, et al. (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313:523-526.
95. Neel C, Etienne L, Li Y, et al. Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. J Virol 84:1464-1476.
96. Van Heuverswyn F, Li Y, Neel C, et al. (2006) Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444:164.
97. Takehisa J, Kraus MH, Ayouba A, et al. (2009) Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J Virol 83:1635-1648.
98. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
99. Subbramanian RA, Cohen EA (1994) Molecular biology of the human immunodeficiency virus accessory proteins. J Virol 68:6831-6835.
100. Trono D (1995) HIV accessory proteins: leading roles for the supporting cast. Cell 82: 189-192.
101. Evans DT, Desrosiers RC (2001) Immune evasion strategies of the primate lentiviruses. Immunol Rev 183:141-158.
102. Bernstein HB, Tucker SP, Hunter E, et al. (1994) Human immunodeficiency virus type 1 envelope glycoprotein is modified by O-linked oligosaccharides. J Virol 68:463-468.
103. Leonard CK, Spellman MW, Riddle L, et al. (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J Biol Chem 265:10373-10382.
104. Kwong PD, Wyatt R, Robinson J, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648-659.
105. Connor RI, Sheridan KE, Ceradini D, et al. (1997) Change in coreceptor use coreceptor use correlates with disease progression in HIV-1-infected individuals. J Exp Med 185:621-628.
106. Cohen GB, Gandhi RT, Davis DM, et al. (1999) The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity 10:661-671.
107. Fauci AS, Mavilio D, Kottilil S (2005) NK cells in HIV infection: paradigm for protection or targets for ambush. Nat Rev Immunol 5:835-843.
108. Mavilio D, Benjamin J, Daucher M, et al. (2003) Natural killer cells in HIV-1 infection: dichotomous effects of viremia on inhibitory and activating receptors and their functional correlates. Proc Natl Acad Sci U S A 100:15011-15016.

109. Martin MP, Gao X, Lee JH, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet 31:429-434.
110. Coffin JM (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267:483-489.
111. Albert J, Abrahamsson B, Nagy K, et al. (1990) Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. Aids 4:107-112.
112. Tremblay M, Wainberg MA (1990) Neutralization of multiple HIV-1 isolates from a single subject by autologous sequential sera. J Infect Dis 162:735-737.
113. di Marzo Veronese F, Reitz MS, Jr., Gupta G, et al. (1993) Loss of a neutralizing epitope by a spontaneous point mutation in the V3 loop of HIV-1 isolated from an infected laboratory worker. J Biol Chem 268:25894-25901.
114. Thali M, Charles M, Furman C, et al. (1994) Resistance to neutralization by broadly reactive antibodies to the human immunodeficiency virus type 1 gp120 glycoprotein conferred by a gp41 amino acid change. J Virol 68:674-680.
115. Watkins BA, Buge S, Aldrich K, et al. (1996) Resistance of human immunodeficiency virus type 1 to neutralization by natural antisera occurs through single amino acid substitutions that cause changes in antibody binding at multiple sites. J Virol 70:8431-8437.
116. Borrow P, Lewicki H, Wei X, et al. (1997) Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat Med 3:205-211.
117. Goulder PJ, Phillips RE, Colbert RA, et al. (1997) Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 3:212-217.
118. Phillips RE, Rowland-Jones S, Nixon DF, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 354:453-459.
119. Couillin I, Culmann-Penciolelli B, Gomard E, et al. (1994) Impaired cytotoxic T lymphocyte recognition due to genetic variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein. J Exp Med 180:1129-1134.
120. Evans DT, O'Connor DH, Jing P, et al. (1999) Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. Nat Med 5:1270-1276.
121. Goulder PJ, Watkins DI (2004) HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol 4:630-640.
122. Allen TM, Altfeld M, Yu XG, et al. (2004) Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. J Virol 78:7069-7078.
123. Draenert R, Le Gall S, Pfafferott KJ, et al. (2004) Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. J Exp Med 199:905-915.
124. Yokomaku Y, Miura H, Tomiyama H, et al. (2004) Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. J Virol 78:1324-1332.
125. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713.
126. Deinard A, Kidd K (1999) Evolution of a HOXB6 intergenic region within the great apes and humans. J Hum Evol 36:687-703.

127. Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.
128. Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A 97:11354-11358.
129. Lacap PA, Huntington JD, Luo M, et al. (2008) Associations of human leukocyte antigen DRB with resistance or susceptibility to HIV-1 infection in the Pumwani Sex Worker Cohort. Aids 22:1029-1038.
130. Goulder P, Conlon C, McLntyre K, McMichael A (1996) Identification of a novel human leukocyte antigen A26-restricted epitope in a conserved region of Gag. Aids 10:1441-1443.
131. Kaslow RA, Carrington M, Apple R, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2:405-411.
132. Klein MR, Miedema F (1995) Long-term survivors of HIV-1 infection. Trends Microbiol 3: 386-391.

Chapter 2

Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research

Natasja G. de Groot[1], Nel Otting[1], Rafael Argüello[2], David I. Watkins[3], Gaby G.M. Doxiadis[1], J. Alejandro Madrigal[2], Ronald E. Bontrop[1]

Immunogenetics 51 (2000), 398-409

[1]Department of Comparative Genetics and Refinement, Biomedical Primate Research Centre, Rijswijk, The Netherlands.
[2]Department of Haematology, The Royal Free Hospital, London, UK.
[3]Department of Pathology, Wisconsin Regional Primate Research Center, Madison, USA.

Abstract

Human immunodeficiency virus (HIV) poses a major threat to humankind. And though, like humans, chimpanzees are susceptible to HIV infection, they are considered to be resistant to the development of the acquired immune deficiency syndrome (AIDS). Little is known about major histocompatibility complex (MHC) class I diversity in chimpanzee populations and, moreover, whether qualitative aspects of Patr class I molecules may control resistance to AIDS. To address these questions, we assayed MHC class I diversity in a West African chimpanzee population and in some animals from other subspecies of chimpanzee. Application of different techniques allowed the detection of 17 full-length *Patr-A*, 19 *Patr-B*, and 10 *Patr-C* alleles. All Patr-A alleles cluster only into the *HLA-A1/A3/A11* family, which supports the idea that chimpanzees have experienced a reduction in their repertoire of *A* locus alleles. The *Patr-B* alleles do not cluster in the same lineages as their human equivalents, due to frequent exchange of polymorphic sequence motifs. Furthermore, polymorphic motifs may have been exchanged between *Patr-A* and *Patr-B* loci, resulting in convergence. With regard to evolutionary stability, the *Patr-C* locus is more similar to the *Patr-A* locus than it is to the *Patr-B* locus. Despite the relatively low number of animals analyzed, humans and chimpanzees were ascertained as sharing similar degrees of diversity at the contact residues constituting the B and F pockets in the peptide-binding side of MHC class I molecules. Our results indicate that within a small sample of a West African chimpanzee population, a high degree of Patr class I diversity is encountered. This is in agreement with the fact that chimpanzees display more mitochondrial DNA variation than humans. In addition, population analyses demonstrated that particular Patr-B molecules, with the capacity to bind conserved HIV-1 epitopes, are characterized by high gene frequencies. These findings have important implications for evaluating immune responses in HIV vaccine studies and, more importantly, may help in understanding the relative resistance of chimpanzees to AIDS.

Introduction

Humans and chimpanzees are closely related and share approximately 98.6% similarity at the DNA level. They are also the only species that can be infected with pathogens such as human immunodeficiency virus (HIV), various hepatitis viruses, and *Plasmodium falciparum* (1-4). For that reason, the chimpanzee is considered to represent an important resource for biomedical research. The host defense to viral infections and other intracellular parasites is to a large extent controlled by the gene products of the major histocompatibility complex (MHC) region. Processed viral segments (peptides) are recognized in the context of MHC class I molecules by the T-cell receptor of cytotoxic T cells, and successful recognition may result in the lysis of the infected cell. Some viruses have developed strategies to interfere with MHC class I cell surface expression in order to escape immune recognition.

In that respect, the absence of MHC class I molecules is scanned by natural killer cells and possibly results in the elimination of the infected target (5).
Thus far, more than 100 chimpanzees housed in different centers have been infected with various HIV strains but only one animal has been diagnosed with the acquired immune deficiency syndrome (AIDS) (6). As such, chimpanzees might be considered equivalents of the human long-term nonprogressor population (1). Protective cytotoxic T lymphocyte (CTL) responses are a potential candidate mechanism that may play a key role in protecting humans and chimpanzees from AIDS, thus justifying a detailed study of the chimpanzee MHC class I repertoire.
MHC class I molecules are composed of a heavy chain which complexes with β_2-microglobulin, and the classical MHC class I molecules are characterized by an extremely high degree of polymorphism (7). In humans, three types of classical *Mhc* class I genes have been defined, designated *HLA-A*, *HLA-B*, and *HLA-C* (8). The evolutionary equivalents are present in chimpanzees as well and have been named *Patr-A*, *Patr-B*, and *Patr-C* (9-13). The *Patr* class I alleles documented so far were obtained from isolated animals belonging to different subspecies but a thorough population analysis of *Patr* class I alleles and their frequencies has not yet been carried out. The core of the Biomedical Primate Research Centre (BPRC) chimpanzee colony was started with animals imported from Sierra Leone, and subspecies designation has been confirmed by mtDNA variation analysis (14). This colony reflects, at least in part, the genetic makeup of the wild population of West African chimpanzees and represents unique material to investigate the diversity of the *Mhc* class I repertoire. In this context, diversity is defined as the combination of polymorphisms observed at the *Patr-A*, *Patr-B*, and *Patr-C* loci.

Materials and methods

Animals

The 35 founder animals studied were caught wild in Sierra Leone and belong to the West African subspecies (*Pan troglodytes verus*). The BPRC chimpanzee colony was serologically typed for the Patr-A and Patr-B specificities and at the molecular level characterized for the polymorphic *Patr-DR*, *Patr-DQ* and *Patr-DP* class II genes (15-18). All animals and their offspring have been pedigreed based on segregation of serological specificities and molecular-defined *Patr* class II gene polymorphisms (16). Two other chimpanzees, Niko and Noah, form a part of this study and belong to the East African subspecies (*P. t. schweinfurthii*), whereas Anita, Brigitte, Victoria, and Wilma seem to belong to the Central African subspecies (*P. t. troglodytes*). Niko and Noah have been naturally infected with simian immunodeficiency virus derived from a chimpanzee (SIVchz). This natural infection with SIVchz and the opportunity to compare *Mhc* class I alleles of different subspecies of chimpanzees was the rationale for including these animals in this study. Discrimination of the different subspecies of chimpanzees in this panel was checked based on mtDNA variation. DNA was used to amplify

the nucleotide sequences of the mitochondrial D loop, and the nucleotide sequences were compared to the database of Morin and co-workers (1994) to assign the subspecies.

Mhc class I gene characterization by reference strand conformation analysis and nucleotide sequencing

Reference strand conformation analysis (RSCA) was used to type the animals for their *Patr-A*, *Patr-B*, and *Patr-C* genes. RSCA was performed at the Department of Haematology, The Royal Free Hospital, London, as has been described elsewhere (19). For nucleotide sequencing, total RNA was extracted from B-lymphoblastoid cell lines using the Rneasy mini kit (Qiagen) according to the manufacturer's instructions. First-strand cDNA synthesis was established using Moloney murine leukemia virus reverse transcriptase (M-MLV RT, Life Technologies). One microliter of oligo(dT) primer was added to 13 μl total RNA preparation and incubated for 2 min at 70 ˚C and an additional 3 min at room temperature. Then 5 μl of M-MLV RT buffer, 2 μl dithiothreitol, 3 ml deoxyribonucleoside triphosphates (dNTPs; 0.6 mm each dNTP), and 1 μl M-MLV RT was added and incubated for 1 h at 37 ˚C. Briefly, amplification of full-length *Mhc* class I nucleotide sequences was performed using 10 μl of cDNA, 0.8 μm of each primer (HLA-5P2 and HLA-3P2) (20) and 5 units of AmpliTaq (Perkin Elmer) in 30 cycles. Each cycle consisted of 60 s at 94 ˚C, 90 s at 63 ˚C, and 70 s at 72 ˚C, with a final amplification of 5 min at 72 ˚C. For the amplification of the *Patr-C*0203* allele, the primers 5'LPHIII (21) and 3'UTBam (5'-GCGGATCCAGTCCCACACAAGGCAGCTG-3') were used. The PCR products were extracted with phenol/chloroform/isoamylalcohol (ph/cl/iso) and precipitated with ethanol. Both the target DNA and the bacteriophage M13 derivatives mp18 and mp19 (Life Technologies) were digested with the restriction endonucleases *Sal*I and *Hin*dIII, extracted with ph/cl/iso, and precipitated with ethanol. The target DNA and vectors were digested with *Bam*HI and *Hin*dIII when the primers 5'LPHIII and 3'UTBam were used. The target DNA was subsequently ligated into the vectors. Competent XL1-Blue cells were used for transformation. Phages were picked after plaque screening with locus-specific biotine-labelled probes [A locus (sense strand): 5'BiotAGGCTGCAAGCAGTGACAGT3', B locus (sense strand): 5'BiotTCCTCCTGCTGCTCTCGGGG3' and C locus (sense strand): 5'BiotCAAGCGCCAGGCACAGGCT3'], grown in culture and used to prepare single-stranded M13 DNA. Finally, sequencing was performed on the ABI 310 automatic sequencer using the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer) to prepare the samples according to the manufacturer's instructions. Sequencing oligonucleotides (20) were used to sequence the full-length *Mhc* class I cDNA. At least two independent PCR reactions were performed and/or the alleles were confirmed by their presence in different animals.

Phylogenetic analysis

Trees were constructed with the neighbor-joining method from the PAUP* version 4.0b2 for Macintosh, written by David L. Swofford, using Kimura's two-parameter method. The same program was used to define bootstrap values based on 1000 resamplings.

Nomenclature and gene frequency

The *Mhc* class I nucleotide sequences have been named, based upon a previously published proposal (22). For example, the equivalents of *HLA-A* locus products have been designated *Patr-A*. The first two digits identify the lineage, while the last two digits reflect the order in which the alleles were found. The chimpanzee *Mhc* class I lineages are characterized by arbitrarily chosen digits. Thus, human and chimpanzee alleles which are identified by similar names (for instance *HLA-A*0201* and *Patr-A*0201*) do not share any obvious evolutionary relationship. The GenBank accession numbers of the sequences described in this study are listed in Table 1. Gene frequencies were determined for the founder population of West African chimpanzees in this study according to standard procedures (Table 1).

Results and discussion

Comparative analysis of Patr class I typing methods

The available alloantisera characterize eight Patr-A specificities designated A108, A111, A112, A121, A129, A122, A126, and A128 and six Patr-B serotypes named B4, B106, B117, B118, B123, and B125 (15). Based on segregation analyses, chimpanzees appear to have particular Patr-A and Patr-B molecules for which the corresponding typing sera are lacking. These specificities are called blanks and are denoted Patr-A- and Patr-B- (Table 1). Sera that define Patr-C molecules are not described.

A large section of the chimpanzee cohort was typed by RSCA. This technique allowed the detection of 17 *Patr-A*, 19 *Patr-B*, and 10 *Patr-C* gene-related fragments (Table 1). Family studies demonstrated that the RSCA fragments segregate in a Mendelian manner. In the case of serotypes A108, A112, A121, A128, B117, and B123, only one corresponding RSCA fragment was detected (Table 1). All other serotypings have been linked to multiple RSCA fragments. As can be seen, the B125 specificity is associated with seven fragments. In the example of most Patr-A blank specificities, corresponding RSCA fragments were detected. However, in the case of particular Patr-A blank (Marco) and Patr-B blank (Susie) specificities, the corresponding RSCA fragments are absent (Table 1).

Based on dissimilar serotypings/RSCA fragment profiles, a panel of chimpanzees was selected for nucleotide sequence analysis. The panel was composed in such a way that, whenever possible, the same allele was confirmed in distinct animals. This approach allowed the detection of 17 full-length *Patr-A*, 19 *Patr-B*, and 10 *Patr-C* alleles. In the case of *Patr-A* and *Patr-B* gene analyses, one can conclude that the existence of an RSCA fragment correlates with the presence of the corresponding full-length cDNA. This indicates that most of the Patr-A blank specificities represent apparently functional genes encoding bona fide gene products for which typing sera are absent. However, for particular Patr-A and Patr-B blank specificities, no RSCA fragment or corresponding cDNA could be characterized (Table 1). This may indicate that on certain chimpanzee haplotypes the corresponding *Patr-A* or *Patr-B* gene may be absent or has undergone gross alterations.

Table 1. Summary of various *Patr* typing data. The previous designations of the different alleles are listed in *parentheses* after the official designations (ND not determined; *P.t.v. Pan troglodytes verus*, *P.t.s. Pan troglodytes schweinfurthii*, *P.t.t. Pan troglodytes troglodytes*).

Serological specificity Patr-	RSCA-fragment	Nucleotide sequence *Patr-*	Gene Frequency	Reference animal	Subspecies	Genebank accession number for nucleotide sequence data
A108	A5	*A*0301 (*03)*	17.5	Hugo	*P.t.v.*	AF168392
A111	A7	*A*0101 (*01)*	10.5	Hans	*P.t.v.*	AF168393
	A10	*A*0201 (*02)*	1.8	Marja	*P.t.v.*	AF168394
A112	A2	*A*0701 (*07)*	12.3	Sherry	*P.t.v.*	AF168395
A121	A11	*A*1101 (*11)*	1.8	Karin	*P.t.v.*	AF168396
A121 (129)	A6	*A*0601 (*06)*	7.0	Flint	*P.t.v.*	AF168397
	A16	*A*0501 (*05)*	7.0	Diana	*P.t.v.*	AF168398
	A19	*A*0602 (*16)*	1.8	Louise	*P.t.v.*	AF165354
A122	A12	*A*1702 (*15)*	ND	Wilma	*P.t.t.*	AF165353
	A15	*A*1201 (*12)*	ND	Brigitte/Victoria	*P.t.t.*	AF168399
A126	A1	*A*0901 (*09)*	14.0	Marco	*P.t.v.*	AF168400
	A4	*A*0401 (*04)*	8.8	Hugo	*P.t.v.*	AF168401
	A18	*A*1301 (*13)*	ND	Victoria	*P.t.t.*	AF168403
A128	A3	*A*1401 (*14)*	7.0	Sherry/Niko	*P.t.v./P.t.s.*	AF168402
A-	A13	*A*0801 (*08)*	ND	Anita/Niko/Noah	*P.t.t./P.t.s.*	AF168404
	A14	*A*1001 (*10)*	ND	Brigitte	*P.t.t.*	AF168405
	A17	*A*0302 (*17)*	1.8	Flint	*P.t.v.*	AF165355
	-	-	8.8	Marco	*P.t.v.*	
B4	B7	*B*16011 (*22)*	3.5	Yoko	*P.t.v.*	AF165359
	B8	*B*0401 (*04)*	1.8	Hans	*P.t.v.*	AF168406
B106	B6	*B*1301 (*13)*	12.3	Karin	*P.t.v.*	AF168407
	B10	*B*1101 (*11)*	ND	Brigitte/Victoria	*P.t.t.*	AF168408
	B11	*B*1201 (*12)*	ND	Victoria	*P.t.t.*	AF168409
	B17	*B*2801 (*23)*	ND	Niko	*P.t.s.*	AF165360
	B19	*B*1202 (*26)*	ND	Wilma	*P.t.t.*	AF165363
B117	B4	*B*1401 (*14)*	8.8	Wodka	*P.t.v.*	AF168410
B118	B5	*B*0201 (*02)*	15.8	Sherry	*P.t.v.*	AF168411
	B18	*B*2303 (*24)*	ND	Niko	*P.t.s.*	AF165361
	B9	*B*2901 (*27)*	3.5	Erik	*P.t.v.*	AF165364
B123	B2	*B*2402 (*25)*	12.3	Liesbeth	*P.t.v.*	AF165362
B125	B1	*B*0101 (*01)*	33.3	Hans	*P.t.v.*	AF168412
	B3	*B*0301 (*03)*	3.5	Frits	*P.t.v.*	AF168413
	B12	*B*3001 (*28)*	ND	Anita	*P.t.t.*	AF179243
	B13	*B*2501 (*18)*	ND	Noah	*P.t.s.*	AF165356
	B14	*B*2601 (*19)*	ND	Noah	*P.t.s.*	AF165357
	B15	*B*0901 (*09)*	1.8	Louise	*P.t.v.*	AF168414
	B16	*B*2701 (*20)*	ND	Brigitte	*P.t.t.*	AF165358
B-	-	-	3.5	Susie	*P.t.v.*	
	C1	*C*0203*	14.0	Erik	*P.t.v.*	AF179244
	C2	*C*0401*	31.6	Marja	*P.t.v.*	AF165365
	C3	*C*0901*	15.8	Susie/Niko/Wilma	*P.t.v./P.t.s./P.t.t.*	AF165366
	C4	*C*0303*	ND	Brigitte/Noah	*P.t.t./P.t.s.*	AF165367
	C4.1	*C*0301*	1.8	Flint	*P.t.v.*	AF165369
	C5	*C*0601*	22.8	Susie	*P.t.v.*	AF165368
	C6	*C*0502*	3.5	Yoko/Brigitte/Victoria	*P.t.v./P.t.t*	AF165370
	C7	*C*1101*	8.8	Louise/Niko	*P.t.v./P.t.s.*	AF165371
	C8	*C*0501*	1.8	Hans	*P.t.v.*	AF165372
	C9	*C*0903*	ND	Noah	*P.t.s.*	AF165373

Patr-A locus analyses

Within the BPRC founder population of 35 chimpanzees, at least 12 *Patr-A* alleles must have been present (Table 1). The nucleotide sequences of two of these alleles, *Patr-A*0302* and *-A*0602* have not been reported previously. The *Patr-A*1001*, *-A*1201*, *-A*1301*, and the newly described *-A*1702* alleles were detected in the subspecies *P. t. troglodytes*. Some of the *Patr-A* alleles found are shared between different subspecies. This is illustrated by the data obtained for Sherry and Niko (*Patr-A*1401*) as well as for Anita, Niko, and Noah (*Patr-A*0801*). If it is true that the different chimpanzee subspecies were separated about 1.5 million years ago (23), one can conclude that some *Patr-A* alleles have remained genetically stable over this time span.

The deduced polypeptide chains from the alpha 1 and alpha 2 domains illustrate that the 17 *Patr-A* molecules are closely related (Fig. 1). Most of the variation seen at the *Patr-A* alleles may have been caused by point mutations. One example is provided by the *Patr-A*0601* and *-A*0602* sequences which differ only for one nonsynonymous mutation, at codon 70, resulting in an amino acid replacement mapping to a contact residue in the peptide-binding site. A phylogenetic tree constructed based on full-length sequences shows a division of the *Patr-A* alleles into two sublineages, designated a-1 and a-2 (Fig. 2A). As can be seen, all 17 *Patr-A* alleles cluster into only one of the six *HLA-A* locus families, namely the *HLA-A1/A3/A11* lineage, as has been documented earlier for a smaller group of *Patr-A* alleles (13). Approximately 5 million years ago, humans and chimpanzees shared a common ancestor, indicating that the *HLA-A1/A3/A11* lineage has been maintained for more than 5 million years. However, not one of the *Patr-A* alleles is identical to an *HLA-A* locus equivalent. As such, all *HLA-A* and *Patr-A* alleles themselves must have a post-species origin.

In humans, the *HLA-A2* lineage is represented by a high number of alleles (7). Orthologues of *HLA-A2*-like alleles have been detected in gorillas (24) but are absent in chimpanzees. Although their exact radiation is debated, humans and the African great apes shared a common ancestor approximately 6 million years ago. Thus one would expect that the ancestral stock of humans, chimpanzees, and gorillas most likely inherited the evolutionary ancestors of the *HLA-A2* lineage. The absence of *HLA-A2*-like sequences in chimpanzees can be accounted for in two ways. First, one could argue that this absence is due to a small founder population/bottleneck effect. This possibility contradicts the fact that even chimpanzee subspecies display far more mtDNA variation than the human species (14). The second possibility may be that due to an infectious disease, a severe reduction of lineages took place at the *Patr-A* locus. Since the *HLA-A2*-like sequences are also absent in the Bonobo (pygmy chimpanzee: *Pan paniscus*), such a selection event may have occurred more than 2 million years ago. One should take into consideration that only a limited sample of bonobos has been analyzed.

Patr-B locus analyses

Ten different *Patr-B* alleles were detected in the West African chimpanzee population studied, and the unreported alleles are designated *Patr-B*16011*, *-B*2402*, and *-B*2901*. Four undescribed alleles (*Patr-B*2303*, *-B*2501*, *-B*2601* and *-B*2801*) were isolated from animals

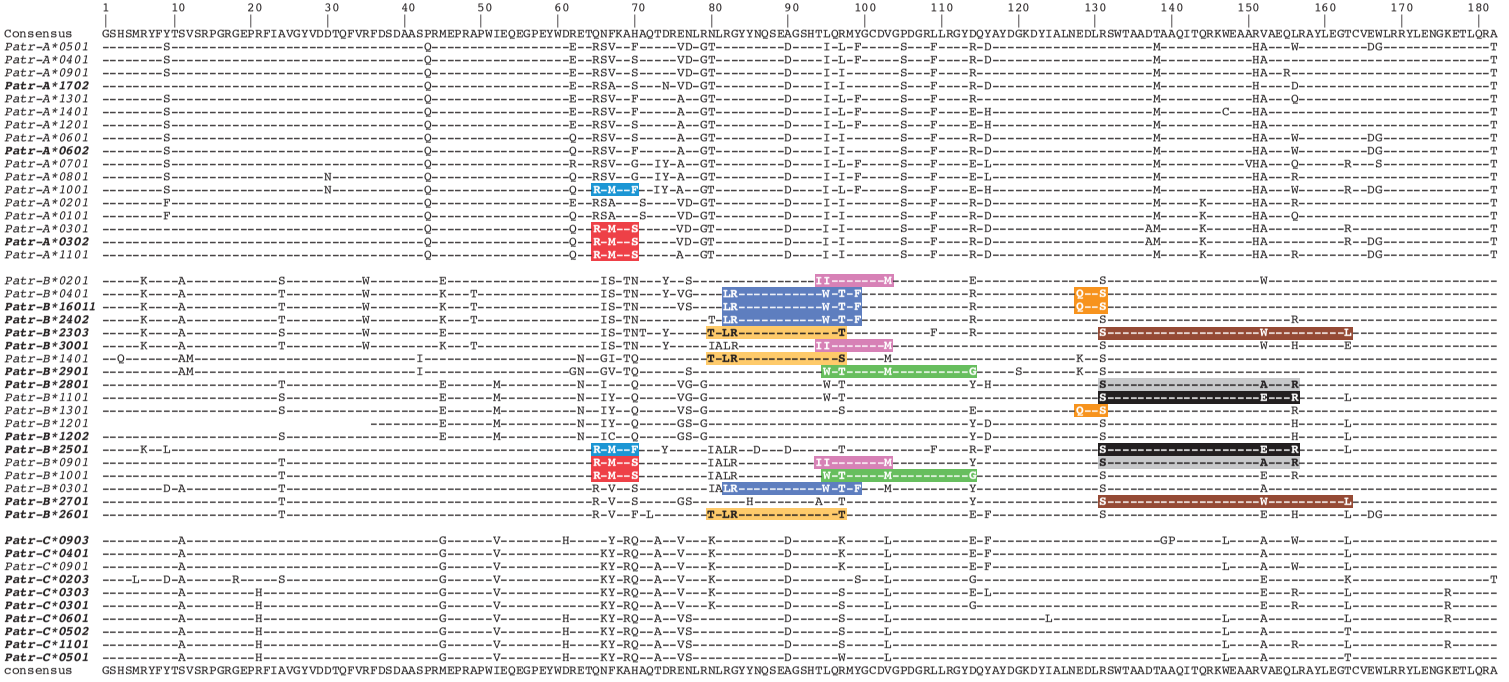


Fig. 1. Deduced alpha 1 and alpha 2 domains of the *Patr* class I nucleotide sequences. Identity to the consensus sequence is indicated by *dashes*, whereas amino acid replacements are depicted by the conventional *one-letter code*. The allele names depicted in *bold* are the previously unreported alleles. The *red* and *blue box* indicate the motifs RMNKAS and RMNKAF, respectively. The other *colored boxes* highlight polymorphic sequence motifs that are shared between distinct Patr-B molecules and may have been subject to recombination.

of the subspecies *P. t. schweinfurthii*, whereas *Patr-B*1101* and *-B*1201* and the newly described *-B*1202*, *-B*2701*, and *-B*3001* alleles are present in *P. t. troglodytes* (Table 1). The alpha 1 and alpha 2 domains of the Patr-B molecules show considerably more amino acid variation than was observed for the Patr-A molecules (Fig. 1). Phylogenetic analysis of full-length *Patr-B* sequences seems to support a division into three major different lineages, designated b-1, b-2, and b-3 (Fig. 2A). When more sequences become available, a further subdivision will probably be likely. Phylogenetic analysis of exon 2 sequences (Fig. 2B) indeed confirmed that the *Patr-B* alleles cluster into three lineages and also show, as previously reported, clustering with particular *HLA-B* equivalents (12). In contrast to the *Patr-A* alleles, different subspecies of chimpanzees do not appear to share identical *Patr-B* alleles, which supports the notion that the *Patr-B* locus is evolving faster than the *Patr-A* locus.

Genetic mechanisms generating polymorphism: evidence for convergence of Patr-A and Patr-B sequences

As was discussed earlier, most variation seen at the *Patr-A* locus appears to result from point mutations. The situation for the *Patr-B* locus seems to be more complex. As has been documented, most variation seen at the *HLA-B* and *Patr-B* loci results from recombination-type events promoting exchange of polymorphic sequence motifs between lineages (12, 25, 26). Our reported sequences support these findings (Fig. 1). Some polymorphisms, however, are shared not only between lineages but also between loci. The two polymorphic amino acid motifs, RNMKAS and RNMKAF (position 65–70), are shared between Patr-A and -B molecules (Fig. 1). For example, the first motif is present in the Patr-A*0302 and -B*0901 molecules of the subspecies *P. t. verus*, whereas the second motif is present in the Patr-A*1001 and -B*2501 molecules of the subspecies *P. t. schweinfurthii* and *P. t. troglodytes* (Fig. 1), respectively. These latter two subspecies are considered to be highly related (23). This sharing of sequence motifs is also reflected by the phylogenetic analyses of the exon 2 sequences (Fig. 2B). As can be seen, the b-3 lineage seems to be more related to the *Patr-A* lineages than to the other *Patr-B* lineages. This effect is not seen, however, when entire *Patr* class I sequences are subjected to phylogenetic clustering (Fig. 2A). More detailed analysis demonstrates that the first motif is identical at the DNA level for the relevant *Patr-A* and *Patr-B* alleles, while the second motif differs only for one synonymous mutation at codon 66 (Fig. 3). There are different mechanisms by which these motifs may have been generated. For instance, the *Patr-A*0401* allele requires only two point mutations to convert its sequence into the motif that is present in *Patr-A*0302* (Fig. 3). The alternative explanation is that the motif has been replaced by a recombination event. A silent mutation, shared between *Patr-A* and *Patr-B* alleles with the RNMKAS motif (Fig. 3), favors the latter possibility because synonymous mutations are considered to evolve under neutral selection. However, it is difficult to determine whether the motif has been moved from *Patr-A* to *Patr-B*, or vice versa. On the contrary, the de novo generation of polymorphic motifs by convergent evolution has been documented for unrelated *Mhc-DR* genes (27, 28). This possibility may be ruled out,

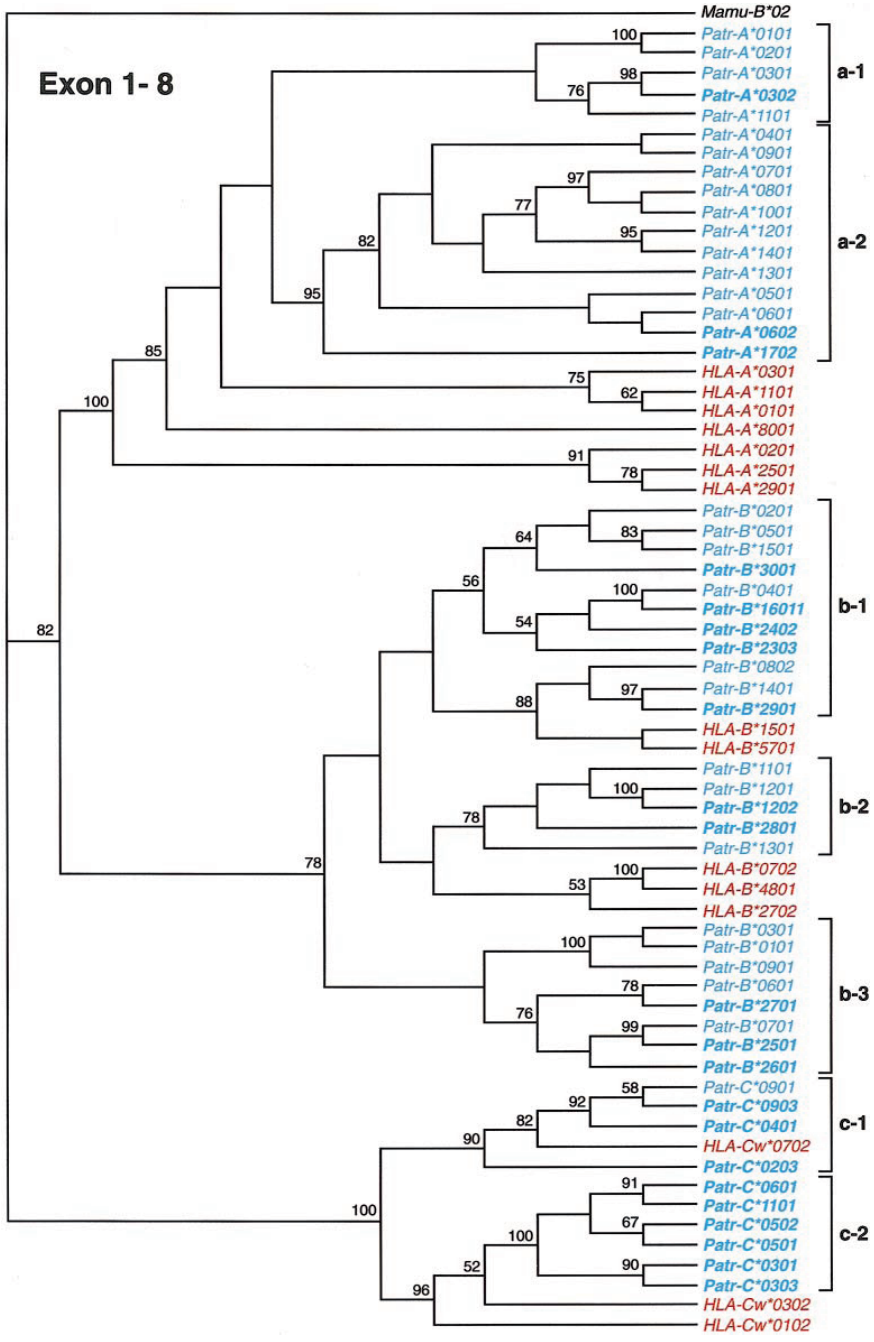


Fig. 2A. Phylogenetic tree of exons 1- 8 of the *Patr-A*, *Patr-B*, and *Patr-C* alleles (*blue*) compared to their *HLA* equivalents (*red*). The brackets illustrate the division of the *Patr-A*, *Patr-B*, and *Patr-C* alleles into the different lineages. The trees are rooted by using the *Mamu-B*02* allele. The allele names depicted in *bold* are the previously unreported alleles. Relevant bootstrap values based on 1000 resamplings are indicated.

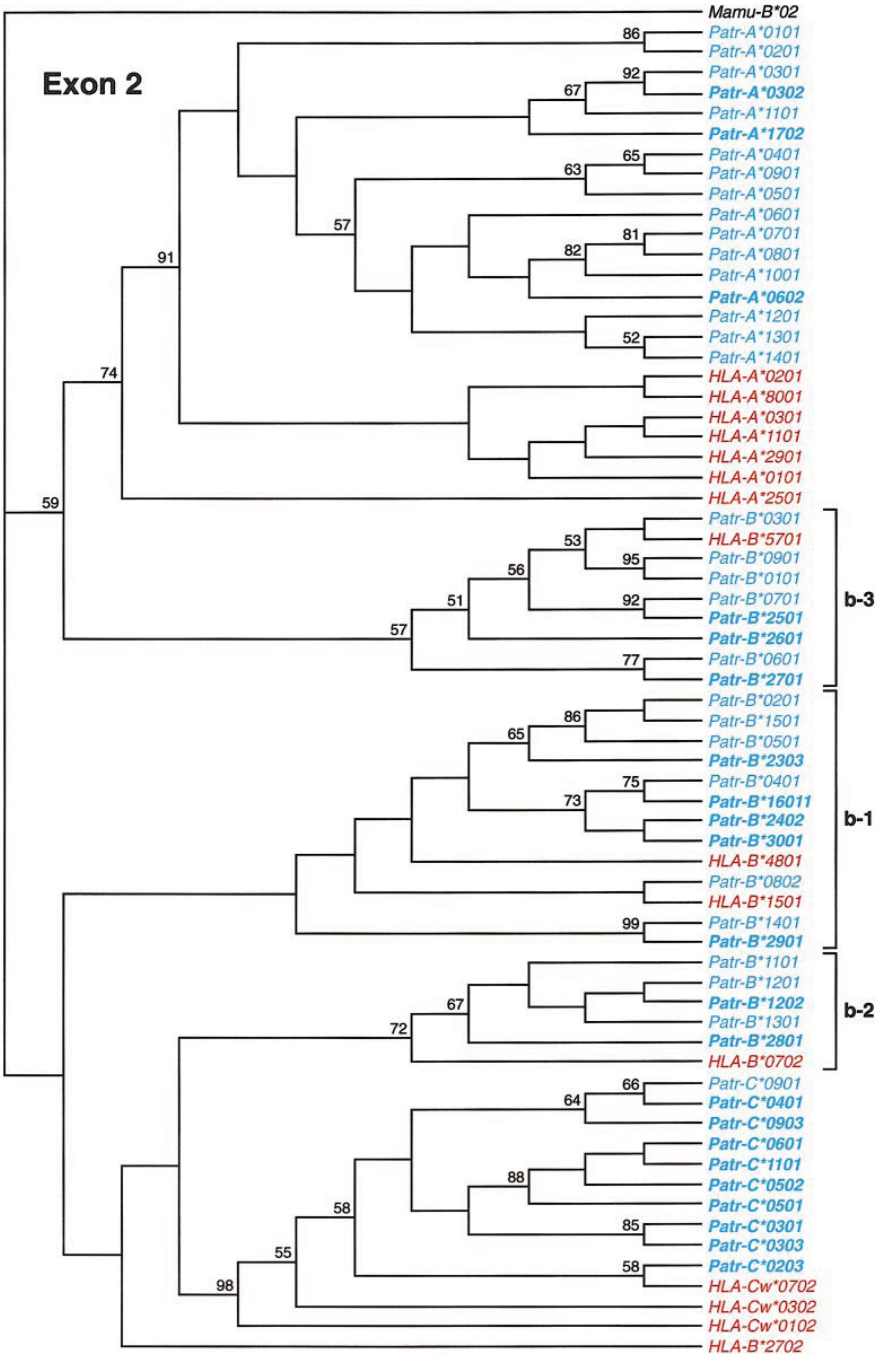


Fig. 2B. Phylogenetic tree of exon 2 of the *Patr-A*, *Patr-B*, and *Patr-C* alleles (*blue*) compared to their human orthologues (*red*). The *brackets* illustrate the division of the *Patr-B* alleles into the different lineages and their clustering with *HLA-B* equivalents.

for the MHC class I motifs discussed, based on codon usage. The polymorphic motifs, RNMKAS and RNMKAF, are part of the peptide-binding site of the MHC class I molecule. Sections of these motifs are involved in the constitution of the B pocket (29), which is known to accommodate anchor residues. As such, the motifs may have important consequences for the selection of peptides that can be bound by these particular MHC molecules. In the human population convergence may have taken place, as highly similar polymorphic epitopes are embedded, for instance, in the HLA-A*0101 and -B*5701 molecules.

Patr-C locus analyses

Only eight *Patr-C* alleles are present in the population of West African chimpanzees studied (Table 1). The *Patr-C*0903* allele was only detected in an animal of the subspecies *P. t. schweinfurthii* (Table 1). Apart from the *Patr-C*0901* allele, none of the *Patr-C* nucleotide sequences had been previously reported or could be retrieved from a databank. The designations *Patr-C*0301*, *-C*0303*, *-C*0401*, *-C*0501*, *-C*0601*, *-C*0901*, and *-C*1101*, however, had been used previously (9). It has also been reported that different species of chimpanzees may share identical *Patr-C* alleles (9). As can be expected, similar observations have been made at the subspecies level, for the alleles *Patr-C*0303*, *-C*0502*, *-C*0901*, and *-C*1101* (Table 1). Based on the limited degree of polymorphism observed in the alpha 1 and 2 domains, the *Patr-C* molecules are more similar to the Patr-A than to the Patr-B molecules (Fig. 1). This observation is supported by phylogenetic analyses showing the division of the *Patr-C* alleles into at least two lineages, designated c-1 and c-2 (Fig. 2A). As reported earlier, the alleles *Patr-C*0401* and *Patr-C*0901* cluster with a representative of the *HLA-Cw*07* lineage (Fig. 2A). The newly described *Patr-C*0203* allele clusters within the same lineage, although it seems to outgroup. When more sequence data become available, it will probably become evident that this allele appears to represent a member of a distinct lineage (30). Members of the *HLA-Cw*07* lineage appear in about 40% of the human population (9). The gene frequency (Table 1) shows that the chimpanzee equivalents, *Patr-C*0401* and *Patr-C*0901*, appear to be present in approximately 47% of the West African chimpanzee population tested. This implies that some *C* allelic lineages have also been under conservative selection for a long time span.

The repertoire of B and F pockets in human and chimpanzee Mhc class I molecules.

Within the chimpanzee founder population of 35 animals, 12 *Patr-A*, 10 *Patr-B*, and 8 *Patr-C* alleles are present, whereas in the human population 87 *HLA-A*, 185 *HLA-B*, and 45 *HLA-C* alleles have been detected (Table 2). Up to now, possibly more than a million people from different ethnic populations have been analyzed for *HLA* class I diversity, whereas only a minute sample of chimpanzees has been tested. In this light, a relatively high number of *Patr* class I alleles were detected in a limited sample of chimpanzees. This suggests that within the entire chimpanzee population, with regard to the number of alleles, most likely as much or even more *Mhc* class I diversity is present than in the human population. Similar observations have been made involving comparisons between humans and

	61	62	63	64	65	66	67	68	69	70	71
	D	R	E	T	Q	N	F	K	A	H	A
consensus	GAC	CGG	GAG	ACA	CAG	AAC	TTC	AAG	GCC	CAC	GCA
*Patr-A*0101*	---	-A-	---	---	-G-	-GT	GCG	---	---	---	T--
*Patr-A*0201*	---	GA-	---	---	-G-	-GT	GCG	---	---	---	T--
*Patr-A*0401*	---	GA-	---	---	-G-	-GT	G-G	---	---	TC-	---
*Patr-A*0501*	---	GA-	---	---	-G-	-GT	G-G	---	---	TT-	---
*Patr-A*0601*	---	-A-	---	---	-G-	-GT	G-G	---	---	TC-	---
*Patr-A*0701*	---	---	---	---	-G-	-GT	G-G	---	---	GG-	---
*Patr-A*0801*	---	-A-	---	--c	-G-	-G-	G--	---	---	GG-	---
*Patr-A*1001*	---	-A-	---	---	-G-	--t	A-G	---	---	TT-	---
*Patr-A*0302*	---	-A-	---	---	-G-	--t	A-G	---	---	TC-	---
*Patr-A*0602*	---	-A-	---	---	-G-	-GT	G-G	---	---	TT-	---
*Patr-A*1702*	---	GA-	---	---	-G-	-GT	GCG	---	---	TC-	---
*Patr-B*0201*	---	--a	---	---	---	-T-	-C-	---	A--	A--	---
*Patr-B*0301*	---	---	---	---	-G-	---	G-G	---	---	TC-	---
*Patr-B*0901*	---	---	---	---	-G-	--t	A-G	---	---	TC-	---
*Patr-B*1101*	---	---	A-C	---	---	-T-	-A-	---	---	--G	---
*Patr-B*1401*	---	---	A-C	---	---	GG-	A--	---	A-G	--G	---
*Patr-B*1202*	---	---	A-C	---	---	-T-	-G-	---	---	--G	---
*Patr-B*2303*	---	--a	---	---	---	-T-	-C-	---	A--	A--	A--
*Patr-B*2501*	---	---	---	---	-G-	---	A-G	---	---	TT-	---
*Patr-B*2601*	---	---	---	---	-G-	---	G-G	---	---	TT-	---
*Patr-B*2801*	---	---	A-C	---	---	-T-	---	---	---	--G	---
*Patr-B*2901*	---	G--	A-C	---	---	GG-	G--	---	A-G	--G	---
*Patr-C*0401*	---	---	---	---	---	--G	-A-	---	CG-	--G	---
*Patr-C*0601*	C--	---	---	---	---	--G	-A-	---	CG-	--G	---
*Patr-C*0903*	C--	---	---	---	---	---	-A-	---	CG-	--G	---
*Patr-C*1101*	C--	---	---	---	--A	--G	-A-	---	CG-	--G	---
*Patr-C*0203*	---	--a	---	---	---	--G	-A-	---	CG-	--G	---

Fig. 3. Different DNA motifs for the *Patr-A*, *Patr-B*, and *Patr-C* alleles at codons which correspond to the amino acid residue stretch 61–71. Identity to the consensus is indicated by *dashes*. Nonsynonymous and synonymous mutations are depicted in *uppercase* and *lowercase* characters, respectively. The *red* and *blue* boxes indicate the coding region for the motifs RMNKAS and RMNKAF, respectively.

chimpanzees with regard to mtDNA variation (23).

The peptide-binding site of the MHC class I molecules comprises six distinct pockets designated A, B, C, D, E, and F. MHC class I molecules generally bind nonamer peptides. The second amino acid residue (numbered from the N-terminus, of a bound peptide) is fixed by the anchor pocket B, whereas the carboxyl-terminus residue of the peptide is bound to pocket F (31). The polymorphic residues within these anchoring B and F pockets are known to make physical contact with the peptide. We have investigated whether human and chimpanzee MHC class I molecules share similar or identical B and F pockets (Table 2). As can be seen, both species share comparable degrees of variation at the polymorphic residues within the peptide-binding site. This would indicate that humans and chimpanzees, in theory, have the possibility to encode identical pockets and thus may have the capacity to bind the same peptides. The question arises whether identical pockets are actually formed. Both observed B pockets in the Patr-C molecules have human equivalents. In the case of the HLA-B and Patr-B molecules, only a few B pockets are shared, whereas such sharing is not observed for

Table 2. Comparison between the West African chimpanzee population and human alleles with regard to the amino acid residues constituting the B and F pocket.

Locus	Number of alleles analyzed	Number of pockets found in the alleles analyzed (B/F)	Number of pockets that can be formed in theory (B/F)	B-pocket residues								F-pocket residues									
				7	9	2	3	4	6	6	6	7	9	7	7	8	8	8	1	1	1	1	1
						4	4	5	3	6	7	0	9	3	7	0	1	4	1	1	2	4	4
																			6	8	3	3	6
Patr-A	12	7/5	144/32	Y	Y	A	V	M	E	N	V	H	Y	T	N	T	L	Y	Y	Y	Y	T	K
										S	M	F	F	I	D				D				C
											A	S							H				
												G							L				
HLA-A	87	23/10	288/96	Y	Y	A	V	M	E	N	V	H	Y	T	N	T	L	Y	Y	Y	Y	T	K
					F				N	K	M	Q	F	I	D	I	A		D				
					S				Q				C		S				H				
					T														N				
Patr-B	10	7/4	3240/18	Y	Y	A	V	M	E	N	V	Q	Y	T	N	T	L	Y	Y	Y	Y	T	K
					D	S		E	N	I	M	N	F		S	I	A						
						T		K		G	S	S			G	N							
											Y												
											I												
HLA-B	185	40/22	16200/576	Y	Y	A	V	M	E	N	M	Q	Y	T	N	T	L	Y	Y	Y	Y	T	K
					D	S		E	N	I	S	N	F		D	I	A		D		S		L
					H	T		K		K	Y	S	S		S	N			H				
								T			C	K			G				L				
								G			F								F				
																			S				
Patr-C	8	2/4	8/16	Y	Y	A	V	G	E	K	Y	Q	Y	A	N	N	L	Y	Y	Y	Y	T	K
					D	S							S		S	K			F				L
HLA-C	45	9/14	64/128	Y	Y	A	V	G	E	K	Y	Q	Y	A	N	N	L	Y	Y	Y	Y	T	K
					D	S				N			F	T	S	K			F				L
					F								S						L				
					S								C						S				

the molecules encoded by the *A* locus. With regard to the F pocket encoded by the *HLA-C* and *Patr-C* locus, only one example is found for sharing. For the Patr-A and Patr-B molecules, only three and two identical F pockets were observed, respectively, that are also present in the human species. This illustrates that there is only a small chance that a particular set of HLA and Patr class I molecules share the same B and F pockets.

The next question that needs to be addressed is how many of the theoretically possible pockets are actually formed. One can calculate the number of pockets that can be formed based on the degrees of freedom of the polymorphic amino acid residues constituting the pockets. The number of B/F pockets that can be generated in the Patr-A, Patr-B, and Patr-C molecules is 144/32, 3240/18, and 8/16, respectively. In the present panel of Patr-A, Patr-B, and Patr-C molecules, the number of different B/F pockets actually observed is 7/5, 7/4, and 2/4, respectively (Table 2). As could be expected, the number of distinct pockets present is considerably lower than the number that can be theoretically formed. Similar observations are done for humans. Although humans and chimpanzees in theory can generate the same type of pockets, the majority of the pockets themselves appear to be unique. As a consequence, human and chimpanzee MHC class I molecules, closely related according to evolutionary standards, may not bind the same peptides. Evolutionarily related HLA and Patr class I molecules indeed have been documented to bind different peptides (1, 4). This situation differs drastically for HLA-DR3-like molecules, present in different species, which have been demonstrated to share the capacity to bind the same peptide (32). The main implication of this finding is that peptide-binding studies are needed to provide valid information on the qualitative aspects of the MHC class I repertoire in chimpanzees. Such studies will also illustrate whether the *Patr-A* locus molecules have the capacity to bind a small or a wide array of peptides.

Gene frequencies and heterozygote advantage

The gene frequencies of the *Patr-A*, *Patr-B*, and *Patr-C* alleles were determined for the 35 founder animals (Table 1). As can be seen, the most frequently observed alleles for each locus are *Patr-A*0301*, *Patr-B*0101*, and *Patr-C*0401*. Within our contemporary colony, all alleles that were present in the founder population can still be traced back. Nevertheless, some gene frequencies have shifted, due to the successful inheritance of certain paternal and maternal haplotypes (data not shown). Particular alleles such as *Patr-B*0201*, *-B*0301*, and *-B*1401* have been linked to the successful presentation of conserved HIV-1 epitopes to CTL (1, 33), and some of these are characterized by high gene frequency numbers. For example, the *Patr-B*0201* allele is present in 15.8% of the founder animals.

The founder population was scanned for MHC class I heterozygosity. One animal was found to be homozygous for the *Patr-A*0601* allele and another one for *Patr-C*0601*, whereas two individuals were found to be homozygous for the combination of the *Patr-B*0101* and *C*0401* alleles. At the BPRC, the breeding animals are housed as stable groups and more than 130 chimpanzees have been born in our facilities. Based on the high gene frequency of some alleles, such as *Patr-A*0301*, *-B*0101*, and *-C*0401*, an increase in the number of homozygous descendants would be expected. But this expectation was not observed when the offspring cohort was analyzed for homozygosity. In this cohort of 92 animals, analyzed by molecular techniques, 17 animals appeared to be homozygous for one, two, or three *Mhc* class I loci. So the majority of the animals seem to be heterozygous. Heterozygous advantage is thought to play an important role in protecting individuals from viral infections such as HIV-1 (34).

Chimpanzees and susceptibility to AIDS

Some reports have suggested that the chimpanzee is the natural reservoir for HIV-1 (35). In this light, the natural resistance of chimpanzees to develop AIDS-like disease is of considerable interest. One plausible explanation for resistance may be that the contemporary chimpanzees mount extremely effective MHC class I-restricted CTL responses to HIV-1-infected cells. Successful lysis of infected cells may help in keeping the viral load down. HIV-1 reverse transcriptase is known to have a high error rate and effective lysis of infected cells would control viral replication and as such minimize the production of mutants that potentially could escape immune recognition. Thus, a detailed characterization of the chimpanzee MHC class I repertoire will facilitate determining the restriction elements and epitopes that may be important in controlling HIV-mediated (1, 33) or hepatitis C virus-mediated (3, 36, 37) diseases.

At the level of mtDNA and some other polymorphic nuclear genes, chimpanzees display far more variation than humans (13, 14, 23). In contrast, chimpanzees seem to have experienced a genetic condensation at the MHC repertoire level. For instance, all known *Patr-A* alleles cluster into only one of the six families of *HLA-A* locus alleles (13). Similar findings have been made for some class II loci as well (16). These condensations are shared by the different species of chimpanzees, and the genetic bottleneck event that affected the MHC repertoire must have taken place more than 2 million years ago. In contrast, with regard to allele number, the *Patr-A*, *Patr-B*, and *Patr-C* loci present within this small chimpanzee population possesses a relative high degree of MHC class I diversity. As such, the chimpanzee populations in general may exhibit at least as much variation as within the total of human populations studied.

Hence, it is tempting to speculate that a retroviral infection in the distant past drastically reduced the numbers of chimpanzees extant at the time, and that the contemporary population represents the offspring of the survivors of this epidemic. The relics of the epidemic are at present still evident in a reduction at the repertoire level of some *Patr* class I and class II lineages. In further support of this hypothesis is the observation that some Patr-B molecules binding epitopes conserved over different HIV-1 clades are characterized by high-frequency numbers (1). The thorough characterization of MHC alleles, restriction elements, and corresponding peptides will be crucial in evaluating the truth of this hypothesis. These types of investigations will help clarify the correlates of protection to virus-mediated diseases, and thus may contribute to the design of future vaccines intended to protect the human population.

Acknowledgements

We would like to thank Ms. Donna Devine for editing the manuscript and Henk van Westbroek for the artwork. This work was partly supported by the EU grants B10 CT 960037, BMH4-CT97-2005, and the MACROPA foundation.

References

1. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, et al. (1999) Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol 162: 2308-2314.
2. Bottius E, BenMohamed L, Brahimi K, et al. (1996) A novel *Plasmodium falciparum* sporozoite and liver stage antigen (SALSA) defines major B, T helper, and CTL epitopes. J Immunol 156:2874-2884.
3. Cooper S, Kowalski H, Erickson AL, et al. (1996) The presentation of a hepatitis C viral peptide by distinct major histocompatibility complex class I allotypes from two chimpanzee species. J Exp Med 183:663-668.
4. Kowalski H, Erickson AL, Cooper S, et al. (1996) Patr-A and B, the orthologues of HLA-A and B, present hepatitis C virus epitopes to CD8+ cytotoxic T cells from two chronically infected chimpanzees. J Exp Med 183:1761-1775.
5. Lanier LL (1997) Natural killer cells: from no receptors to too many. Immunity 6:371-378.
6. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
7. Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67-74.
8. Bodmer JG, Marsh SG, Albert ED, et al. (1999) Nomenclature for factors of the HLA system, 1998. Tissue Antigens 53:407-446.
9. Cooper S, Adams EJ, Wells RS, et al. (1998) A major histocompatibility complex class I allele shared by two species of chimpanzee. Immunogenetics 47:212-217.
10. Lawlor DA, Ward FE, Ennis PD, et al. (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268-271.
11. Mayer WE, Jonker M, Klein D, et al. (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. Embo J 7:2765-2774.
12. McAdam SN, Boyson JE, Liu X, et al. (1994) A uniquely high level of recombination at the *HLA-B* locus. Proc Natl Acad Sci U S A 91:5893-5897.
13. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
14. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
15. Balner H, van Vreeswijk W, Roger JH, D'Amaro J (1978) The major histocompatibility complex of chimpanzees: identification of several new antigens controlled by the A and B loci of ChLa. Tissue Antigens 12:1-18.
16. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
17. Bontrop RE, Otting N, Slierendregt BL, Lanchbury JS (1995) Evolution of major histocompatibility complex polymorphisms and T-cell receptor diversity in primates. Immunol Rev 143:33-62.
18. Otting N, Doxiadis GG, Versluis L, et al. (1998) Characterization and distribution of *Mhc-DPB1* alleles in chimpanzee and rhesus macaque populations. Hum Immunol 59: 656-664.
19. Arguello JR, Little AM, Pay AL, et al. (1998) Mutation detection and typing of polymorphic loci through double-strand conformation analysis. Nat Genet 18:192-194.

20. Ennis PD, Zemmour J, Salter RD, Parham P (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 U S A 87:2833-2837.
21. Watkins DI, Garber TL, Chen ZW, et al. (1991) Unusually limited nucleotide sequence variation of the expressed major histocompatibility complex class I genes of a New World primate species (*Saguinus oedipus*). Immunogenetics 33:79-89.
22. Klein J, Bontrop RE, Dawkins RL, et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217-219.
23. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
24. Lawlor DA, Warren E, Taylor P, Parham P (1991) Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. J Exp Med 174:1491-1509.
25. Belich MP, Madrigal JA, Hildebrand WH, et al. (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature 357:326-329.
26. Watkins DI, McAdam SN, Liu X, et al. (1992) New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature 357: 329-333.
27. Antunes SG, de Groot NG, Brok H, et al. (1998) The common marmoset: a new world primate species with limited *Mhc* class II variability. Proc Natl Acad Sci U S A 95:11745-11750.
28. Gustafsson K, Germana S, Hirsch F, et al. (1990) Structure of miniature swine class II DRB genes: conservation of hypervariable amino acid residues between distantly related mammalian species. Proc Natl Acad Sci U S A 87:9798-9802.
29. Hashimoto K, Okamura K, Yamaguchi H, et al. (1999) Conservation and diversification of MHC class I and its related molecules in vertebrates. Immunol Rev 167:81-100.
30. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51:410-424.
31. Falk K, Rotzschke O, Stevanovic S, et al. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290-296.
32. Geluk A, Elferink DG, Slierendregt BL, et al. (1993) Evolutionary conservation of major histocompatibility complex-DR/peptide/T cell interactions in primates. J Exp Med 177: 979-987.
33. Balla-Jhagjhoorsingh S, Mooij P, Koopman G, et al. (1999) Differential cytotoxic T-lymphocyte (CTL) responses in HIV-1 immunised sibling chimpanzees with shared MHC haplotypes. Immunol Lett 66:61-67.
34. Carrington M, Nelson GW, Martin MP, et al. (1999) HLA and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. Science 283:1748-1752.
35. Gao F, Bailes E, Robertson DL, et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature 397:436-441.
36. Cooper S, Erickson AL, Adams EJ, et al. (1999) Analysis of a successful immune response against hepatitis C virus. Immunity 10:439-449.
37. Santra S, Fultz PN, Letvin NL (1999) Virus-specific cytotoxic T lymphocytes in human immunodeficiency virus type 1-infected chimpanzees. J Virol 73:7065-7069.

Chapter 3

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

Natasja G. de Groot[1], Nel Otting[1], Gaby G.M. Doxiadis[1], Sunita S. Balla-Jhagjhoorsingh[1], Jonathan L. Heeney[1], Jon J. van Rood[2], Pascal Gagneux[3], and Ronald E. Bontrop[1]

Proceedings of the National Academy of Sciences USA 99 (2002), 11748-11753

[1]Department of Comparative Genetics and Refinement, and Virology, Biomedical Primate Research Centre, Rijswijk, The Netherlands.
[2]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands.
[3]Glycobiology Research and Training Center CMM-East, University of California at San Diego, La Jolla, USA.

Abstract

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.

Introduction

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, 16). 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 (16). 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 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 6950-bp fragment [containing complete exon 2 (270 bp), intron 2 (241 bp), and exon 3 (276 bp), and partly intron 1 (695 bp) and intron 3 (660 bp)], using MHC locus-specific primers: 5AIn/3AIn for the *Mhc-A* locus, 5BIn/3BIn for the *-B* locus, and 5CIn/3BCIn for the *-C* locus (24). PCR (25 μl) contained gDNA (0.5μg), 0.2μM of each primer, 1.5mM $MgCl_2$ (for the *B* locus occasionally a concentration of 1.75 mM $MgCl_2$ 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 (27). 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 (28) and Fu's Fs test (29) 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.

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
HLA-			
A	243	40	15
B	478	33	23
Cw	115	15	12
Patr-			
A	26	18	8
B	43	16	11
C	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.

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. 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 (15, 30, 31). 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

A

Concensus	13 G	14 G	17 C	23 G	25 .	26 .	33 C	38 C	46 C	48 C	51 G	60 A	64 C	65 G	80 G	91 A	109 C	122 C	146 T	149 C	151 C	158 A	159 .	181 C	186 C	187 G	191 T	197 G	198 G	199 G	200 C	201 C	204 G	206 G	212 C	217 G	220 T	230 G	231 C	238 G	239 G	
*HLA-A*01*	-	-	-	-	.	.	-	-	-	-	-	-	-	-	-	-	T	T	-	-	-	-	.	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	a1
*HLA-A*30*	-	-	G	-	.	.	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	.	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	a2
*HLA-A*3002*	-	C	G	-	.	.	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	.	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	
*HLA-A*03011*	-	-	G	-	.	.	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	a3
*HLA-A*74*	-	-	G	-	.	.	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	T	-	-	
*HLA-A*29*	-	-	-	-	.	.	-	T	-	-	-	G	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	T	-	-	
*HLA-A*31*	A	-	-	-	.	.	-	T	-	-	-	-	-	A	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	T	-	-	a4
*HLA-A*32*	-	-	-	-	.	.	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	T	-	-	
*HLA-A*34*	-	-	-	-	.	.	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	T	-	-	
*HLA-A*11*	-	-	-	-	.	.	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	-	A	T	-	-	-	-	-	a5
*HLA-A*23*	-	-	-	-	.	.	-	-	-	-	-	G	-	-	-	-	-	-	A	-	-	-	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-/A	a6
*HLA-A*02*	-	-	-	-	.	.	-	T	-	-	-	-	-	-	-	G	-	-	-	-	-	-	.	-	-	A	-	-	-	-	-	G	-	-	-	-	C	-	-	-	
*HLA-A*69*	-	-	-	-	.	.	-	T	-	-	-	-	-	-	-	G	-	-	-	-	-	-	.	-	-/A	-	-	-	-	-	-	G	-	-	-	-	C	-	-	C	-	a7
*HLA-A*25*	-	-	-	-	.	.	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	G	-	-	-	-	C	T	-	C	-	
*HLA-A*80*	-	-	-	-	.	.	-	-	-	T	-	-	-	-	-	-	-	-	-	G	-	-	.	-	-	-	G	-	-	-	-	G	-	-	-	-	C	-	-	-	-	a8
*Patr-A*0101*	-	-	-	-	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	T	-	-	-	x	x	x	x	x	-	-	-	-	C	-	-	-	-	a2
*Patr-A*0301*	-	-	-	-	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	.	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	a5
*Patr-A*0401*	-	-	-	-	.	.	-	-	T	-	C	G	-	-	-	-	-	-	A	T	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	
*Patr-A*0701*	-	-	-	A	.	.	-	-	-	-	C	G	-	-	-	-	-	-	A	T	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	
*Patr-A*1001*	-	-	-	-	.	.	-	-	-	-	C	G	-	-	-	-	-	-	A	T	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	a9
*Patr-A*1801*	-	-	-	-	G	C	-	-	-	-	C	G	-	-	-	-	-	-	A	T	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	
*Patr-A*0801*	-	-	-	-	.	.	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	T	-	-	-	-	A	A	-	-	-	
*Patr-A*1702*	-	-	-	-	.	.	-	-	-	-	C	G	-	-	A	-	-	-	A	-	-	G	.	-	-	C	-	-	-	-	-	-	-	-	-	-	A	-	-	-	a10

B

concensus	14 G	17 C	24 C	26 G	35 T	38 C	40 .	41 .	42 .	44 T	49 C	53 T	58 C	63 G	66 G	69 C	70 C	71 G	78 C	79 G	84 C	89 T	92 G	93 C	95 C	97 C	98 C	99 T	103 G	107 C	120 G	122 C	132 G	133 C	142 C	148 C	149 G	177 C	198 T	215 .	216 .	217 .	218 .	219 .	227 .	229 C	230 G	232 G	243 G	246 C	248 G	
*HLA-B*4701*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	x	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	T	
*HLA-B*4901*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	T	
*HLA-B*5001*	-	-	-/G	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x/-	T/-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	
*HLA-B*4501*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	T	
*HLA-B*27052*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	G	-	
*HLA-B*1801*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	-	
*HLA-B*5701*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-/A	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	-	
*HLA-B*4402*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	-	
*HLA-B*1501*	-	T	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	T	b1
*HLA-B*1517*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	T	
*HLA-B*0702*	-	-	-	-	-	-	.	.	.	-	-	-	A/-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	T	-	-	-	-	
*HLA-B*1302*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	-	
*HLA-B*1513*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	-	
*HLA-B*3801*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	-	
*HLA-B*39062*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	T	-	-	-	
*HLA-B*1559*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	-	-	T	
*HLA-B*0801*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	-	
*HLA-B*4101*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	T	
*HLA-B*4201*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	G	-	-	-	A	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	-	-	b2
*HLA-B*4001*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	T	-	-	-	-	b3
*HLA-B*1401*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	-	-	-	-	G	-	b4
*HLA-B*1402*	-	-	-	-	-	-	.	.	.	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	G	-	.	.	.	.	.	G	-	-	T	-	G	-	
*HLA-B*7301*	-	T	-	-	C	T	C	C	C	A	G	-	-	-	T	T	-	-	T	A	-	-	-	-	-	-	x	x	x	-	-	-	A	C	A	-	A	C	T	-	.	.	.	.	.	G/.	-	T	-	A	-/G	T	b6
*Patr-B*0201*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	T	-	-	G	T	
*Patr-B*1201*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	T	-	-	G	-	
*Patr-B*2501/*2601*	T	-	-	-	-	-	.	.	.	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	T	-	-	G	-	
*Patr-B*1101*	-	-	-	A	-	-	.	.	.	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	T	-	-	G	-	b3
*Patr-B*1301*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	T	-	-	G	-	
*Patr-B*0401*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	.	.	.	.	.	-	T	-	-	G	-	
*Patr-B*1401*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	G	T	-	-	G	-	
*Patr-B*0802*	-	-	-	-	-	-	.	.	.	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	T	-	-	G	-	
*Patr-B*0101*	A	-	-	-	-	-	.	.	.	-	-	G	G	-	-	-	A	C	-	-	G	-	-	-	-	-	x	x	x	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	T	-	-	G	-	b5
*Patr-B*0901*	A	-	-	-	-	-	.	.	.	-	-	G	G	-	-	-	A	C	-	-	G	-	-	-	-	-	x	x	x	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	G	T	
*Patr-B*3001*	-	-	-	-	-	-	.	.	.	-	-	G	G	-	-	-	A	C	-	-	G	-	-	-	-	-	x	x	x	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	-	-	-	-	G	T	

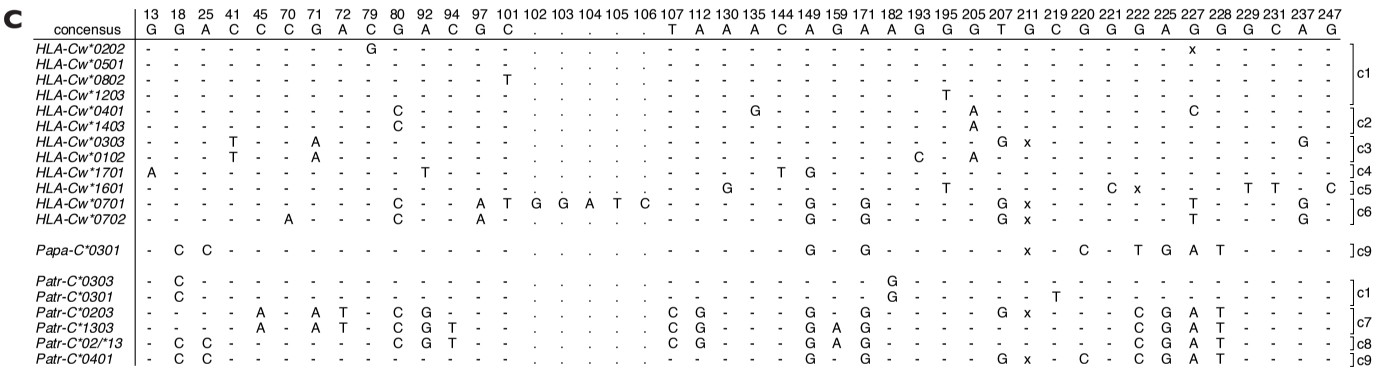


Fig. 1. Polymorphic nucleotide positions in *HLA* and *Patr* intron 2 sequences. Identity to the consensus sequence (depicted at the top) is indicated by dashes. Substitutions and inserts are depicted by the conventional one-letter code, deletions are marked by "x". For instance, "-/A" indicates that 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).

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

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

The table list, 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.

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).

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

		Species	
Mhc-locus	Intron 2 lineage	*HLA*	*Patr*
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.

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 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 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; (28)), 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 Fs test (29) 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). 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

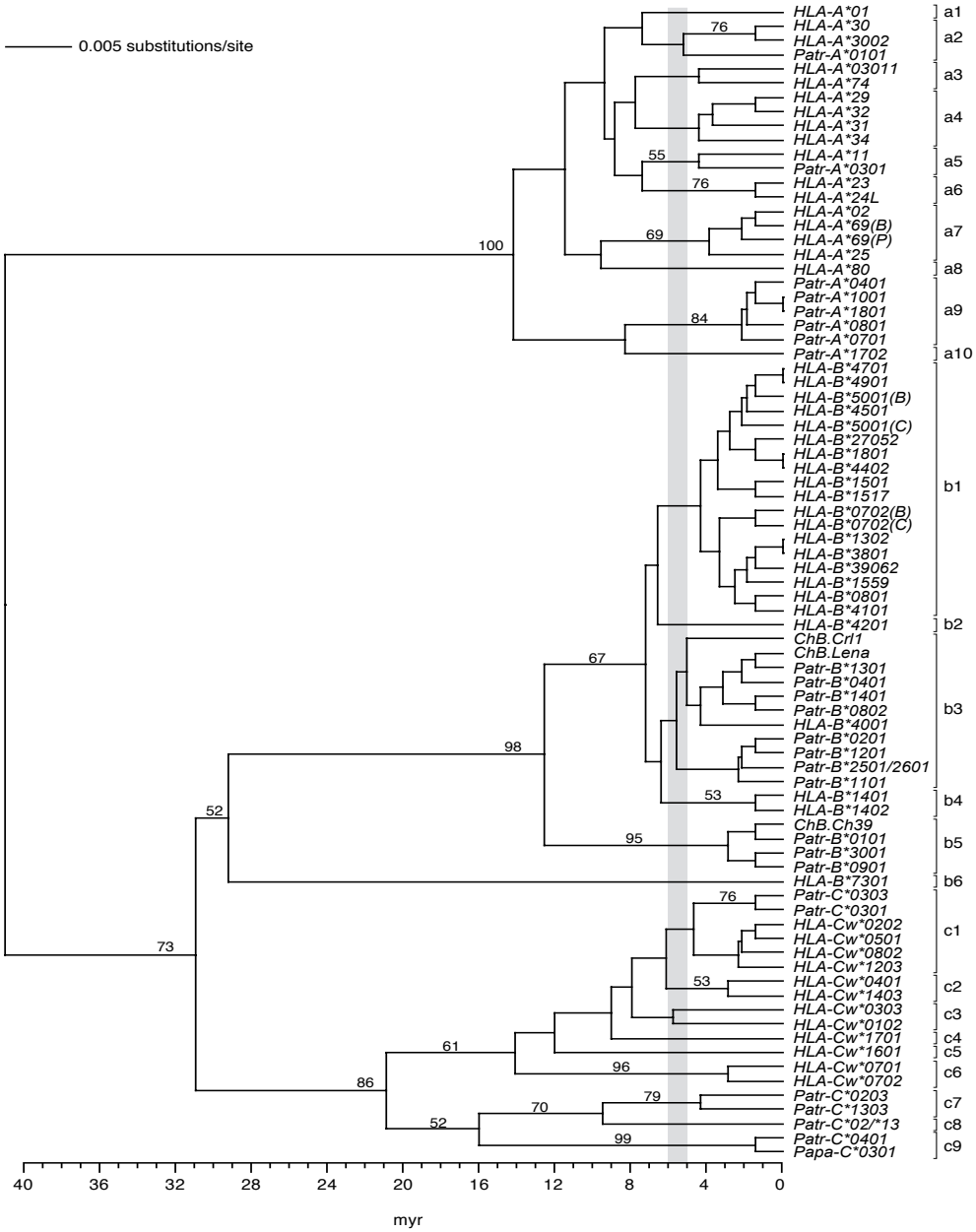


Fig. 2. UPGMA phylogenetic tree of human and chimpanzee MHC class I intron 2 sequences. The divergence time was estimated based on a substitution rate for introns of $1.4*10^{-9}$ per site per year (33). The bar between 5 and 6 myr highlights the period when humans and chimpanzees shared a common ancestor. 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.

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 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 (32) and exon (16) 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 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)

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

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

The *P* value is calculated for Tajima's D.

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 (33-35).

Natural infections with SIVcpz, the closest relative of HIV-1, have been documented in at least six chimpanzees (36) and one free-ranging wild chimpanzee (37). 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 (38). 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 (39). 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 (40). 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 (33), 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 (41). However, humans and chimpanzees recognize such epitopes in the context of MHC class I molecules that group into distinct lineages (33), 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 (42). 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. (37) 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 developing 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.

Acknowledgements

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.

References

1. Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67-74.
2. Bjorkman PJ, Saper MA, Samraoui B, et al. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512-518.
3. Carrington M, Nelson GW, Martin MP, et al. (1999) HLA and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. Science 283:1748-1752.
4. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
5. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
6. Gagneux P, Gonder MK, Goldberg TL, Morin PA (2001) Gene flow in wild chimpanzee populations: what genetic data tell us about chimpanzee movement over space and time. Philos Trans R Soc Lond B Biol Sci 356:889-897.
7. Fujiyama A, Watanabe H, Toyoda A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
8. Sibley CG, Ahlquist JE (1987) DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J Mol Evol 26:99-121.
9. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713.
10. Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.
11. Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A 97:11354-11358.
12. Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68:444-456.
13. de Groot NG, Otting N, Arguello R, et al. (2000) Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. Immunogenetics 51:398-409.
14. Lawlor DA, Ward FE, Ennis PD, et al. (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268-271.
15. McAdam SN, Boyson JE, Liu X, et al. (1994) A uniquely high level of recombination at the *HLA-B* locus. Proc Natl Acad Sci U S A 91:5893-5897.
16. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
17. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51:410-424.
18. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
19. Mayer WE, Jonker M, Klein D, et al. (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J 7:2765-2774.
20. Bodmer WF (1972) Evolutionary significance of the HL-A system. Nature 237:139-145.
21. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.

22. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci U S A 86:958-962.
23. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624-626.
24. Cereb N, Maye P, Lee S, et al. (1995) Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. Tissue Antigens 45:1-11.
25. Jukes TH, Cantor CR (1969) Mammalian Protein Metabolism, ed. (H.N. Munro, Academic, New York), pp. 21-32.
26. Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998) Recent origin of HLA-DRB1 alleles and implications for human evolution. Nat Genet 18:237-242.
27. Swofford DL (1992) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), Version 4
28. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.
29. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.
30. Belich MP, Madrigal JA, Hildebrand WH, et al. (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature 357:326-329.
31. Watkins DI, Chen ZW, Hughes AL, et al. (1990) Evolution of the MHC class I genes of a New World primate from ancestral homologues of human non-classical genes. Nature 346:60-63.
32. Cooper S, Adams EJ, Wells RS, et al. (1998) A major histocompatibility complex class I allele shared by two species of chimpanzee. Immunogenetics 47:212-217.
33. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, et al. (1999) Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol 162: 2308-2314.
34. Cooper S, Erickson AL, Adams EJ, et al. (1999) Analysis of a successful immune response against hepatitis C virus. Immunity 10:439-449.
35. Daubersies P, Thomas AW, Millet P, et al. (2000) Protection against Plasmodium falciparum malaria in chimpanzees by immunization with the conserved pre-erythrocytic liver-stage antigen 3. Nat Med 6:1258-1263.
36. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287:607-614.
37. Santiago ML, Rodenburg CM, Kamenya S, et al. (2002) SIVcpz in wild chimpanzees. Science 295:465.
38. Korber B, Muldoon M, Theiler J, et al. (2000) Timing the ancestor of the HIV-1 pandemic strains. Science 288:1789-1796.
39. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
40. Mwaengo DM, Novembre FJ (1998) Molecular cloning and characterization of viruses isolated from chimpanzees with pathogenic human immunodeficiency virus type 1 infections. J Virol 72:8976-8987.
41. Kaslow RA, Carrington M, Apple R, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2:405-411.
42. Bertoni R, Sette A, Sidney J, et al. (1998) Human class I supertypes and CTL repertoires extend to chimpanzees. J Immunol 161:4447-4455.

Chapter 4

Reduced MIC gene repertoire variation in West African chimpanzees as compared to humans

Natasja G. de Groot[1], Christian A. Garcia[2], Ernst J. Verschoor[1], Gaby G.M. Doxiadis[1], Steven G.E. Marsh[2], Nel Otting[1] and Ronald E. Bontrop[1]

Molecular Biology and Evolution 22 (2005), 1375-1385

[1]Department of Comparative Genetics and Refinement, and Virology, Biomedical Primate Research Centre, Rijswijk, The Netherlands
[2]Anthony Nolan Research Institute and Royal Free and UCL School of Medicine, Royal Free Campus, London, UK

Abstract

The human major histocompatibility complex class I chain-related (*MIC*) genes are members of a multicopy family showing similarity to the classical *HLA-A*, *-B*, and *-C* genes. Only the *MICA* and *MICB* genes produce functional transcripts. In chimpanzees, however, only one *MIC* gene is expressed, showing an intermediate character, resulting from a deletion fusing the *MICA* and *MICB* gene segments together. The present population study illustrates that all chimpanzee haplotypes sampled possess the hybrid *MICA/B* gene. In contrast to the human situation this gene displays reduced allelic variation. The observed repertoire reduction of the chimpanzee *MICA/B* gene is in conformity with the severe repertoire condensation documented for *Patr-B* locus lineages, probably due to the close proximity of both genes.

Introduction

The major histocompatibility complex class I chain-related gene (*MIC*), previously also called PERB11, was first described in 1994 (1, 2). The gene is located in the major histocompatibility complex (MHC) region. The MHC, playing an important role in the immune defense against pathogens, is characterized by its extensive degree of polymorphism at the population level. Also, in the relatively small sample size of chimpanzees that have been analyzed so far, a high degree of *Patr* class I diversity, with regard to allele numbers, has been detected (3, 4). At first sight, this agrees with the finding that chimpanzees display more mitochondrial DNA (mtDNA) variation than humans do (5, 6), and similar findings have been reported for particular nuclear genes (7, 8), although a recent publication shows that the variation in chimpanzees is not as high as a magnitude of four but more in the range of 1.5 compared to humans (9). A more detailed analysis of the MHC class I intron 2 sequences in chimpanzees showed that the MHC class I gene repertoire, with respect to the number of lineages, is severely reduced as compared to humans. This repertoire reduction is evident for *Patr-A* (10) but is most prominent for the *Patr-B* locus lineages (11).

Seven *MIC* genes are distinguished in the human genome. *MICA* and *MICB* produce functional transcripts, whereas *MICC* to *MICG* are pseudogenes (12). The *MICA* and *MICB* genes show a high degree of similarity to the classical MHC class I genes but are distinguished by their disparate organization of exons/introns. Moreover, they do not associate with β_2-microglobulin (β_2m), and their expression is not induced by type I/II interferons. Furthermore, the *MICA* and *MICB* genes are predominantly expressed on fibroblast and epithelial cells (12). A part of the *MICA* and *MICB* promoter region shows similarity to heat shock protein gene promoters, and upregulation of the MICA and MICB molecules after heat shock has been reported (13). In conclusion, MICA and MICB seem to play a role in the detection of cell-stress and appear to react preferentially with the ligands Vδ1 γδ TCR and NKG2D to induce an immune response (13-15). In chimpanzees only one functional *MIC* gene

has been described (16), and this gene appears to have an intermediate character as compared to human *MICA* and *MICB* genes (17, 18). Studies on *MIC* gene polymorphism in chimpanzee populations are absent, and only one allele, *Patr-MIC1*, has been reported (16). The human *MICA* and *MICB* genes show a considerable degree of polymorphism, and to date 54 *MICA* and 14 *MICB* alleles have been identified (19).

Compared to humans, chimpanzees have a severely reduced MHC class I gene repertoire caused by an ancient selective sweep (11), and they appear to lack the equivalents of the *HLA-DRB1*04* and *HLA-DRB1*08* lineages (20). The present study was initiated to investigate whether the repertoire reduction, most prominent for the *Patr-B* locus lineages, is only restricted to the MHC class I and II genes or also extends to other genes located in the MHC region. For that reason we investigated the polymorphism at the chimpanzee *MIC* gene, which is located in the direct neighborhood of the *Patr-B* locus (21). In humans, some of the *MICA* alleles and closely linked *HLA-B* alleles show a high degree of linkage disequilibrium (12). Here we report on *MIC* gene variation and linkage disequilibria in a West African chimpanzee population together with data on chimpanzees of other subspecies.

Materials and Methods

Animals

The chimpanzee (*Pan troglodytes*) colony (approximately 100 individuals) at the Biomedical Primate Research Centre (BPRC) started with 35 founder animals originating from Sierra Leone and belonging to the subspecies *Pan troglodytes verus* (West Africa). The animals are characterized at the molecular level for MHC class I and II gene polymorphisms (4, 20). Their offspring have been pedigreed based on segregation of Patr-A and Patr-B serotypes and molecularly defined *Patr* class I and II gene polymorphisms. Three animals of the *Pan troglodytes troglodytes* (Central Africa) and *Pan troglodytes schweinfurthii* (East Africa) subspecies have been included in this study. For the two bonobos (*Pan paniscus*) used in this study, the DNAs were analyzed with 12S primers to determine potential relationships (22).

mtDNA analysis

Genomic DNA (gDNA), obtained from Epstein-Barr virus-transformed B-cell lines, was used to amplify the mitochondrial D-loop (380 bp) sequences. To assign the subspecies, the nucleotide sequences were compared to published sequences (5, 23). The polymerase chain reaction (PCR) mixture (50 µl) contained gDNA (50 ng), 1 µM of the primers (table 1), 1xPCR buffer + bovine serum albumin, 2 mM $MgCl_2$, 0.2 mM of each deoxyribonucleoside triphosphate (dNTP), and 2.5 units (U) *Taq* polymerase Gold (Applied biosystems, Foster City, USA). A touchdown PCR consisting of the following cycles was run: 1 cycle 15 min at 94°C, 2 cycles of 30 s at 94°C, 30 s at 63°C, 30 s at 72°C; after each two cycles the annealing temperature is decreased by 2°C until 55°C was reached,

Table 1. MIC-Specific PCR Primers and Internal Sequencing Primers.

Name of primer	Sequence	Fragment	Primer for
Contr-forw mtDNA	5'CATGGGGAAGCAGATTTGGGTACCAC3'		PCR
Contr-rev mtDNA	5'CACGGAGGATGGTGACCAAGGG3'		PCR
5'MICA intron 1	5'TCTTGTCCCTTTGCCCGTGTGCAT3'	A	PCR
3'MICA intron 3	5'CGATGTGCCAACAGGAAATGCCTT3'	A	PCR
5'MICPatr-ex3	5'AGTCCTCCAGAGCTCAGACCTTGG3'	B	PCR
3'HOSAE5-3	5'CCTTACCATCTCCAGAAACTGC3'	B	PCR
5'MICPatr4-5nw	5'TGAAGGTGAAGGTCCAGGATCTGT3'		Sequencing
3'MICPatr4-5nwrev	5'ACAGATCCTGGACCTTCACCTTCA3'		Sequencing
3'MICPatr4-3nwrev	5'CCAGGGTCACCCCAGGCTCACCA3'		Sequencing
5'MICPatr5-5nw	5'TTTTTTTTTCAGGGAAGGCACTGA3'		Sequencing

followed by 4 cycles of 30 s at 94°C, 30 s at 53°C, 30 s at 72°C, and 25 cycles of 30 s at 94°C, 30 s at 50°C, 30 s at 72°C, with a final extension of 5 min at 72°C. The PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany), and sequenced directly on an ABI 310 automatic sequencer using the ABI Prism dRhodamine Terminator Cycle sequencing Ready Reaction Kit (Applied Biosystems) to prepare the samples according to the manufacturer's instructions. The products were sequenced from the 5' and 3' ends. Sequences reported in this publication have been deposited in the EMBL database (accession numbers AJ863489-AJ863508).

MIC gene characterization by reference strand conformation analysis

Reference strand conformation analysis (RSCA) was used to screen the chimpanzees for *MIC* gene polymorphisms. Locus-specific primers, amplifying human *MICA*, were used to independently amplify exon 2 and 3 in chimpanzees (24). RSCA was performed at the Anthony Nolan Research Institute, London, as has been described in detail (25).

MIC gene characterization by direct nucleotide sequencing

gDNA was used for the PCR amplification. Two independent overlapping PCR fragments were generated, using *MIC*-specific primers (table 1), covering the *MIC* gene from exon 2 to exon 5. This resulted in a 1,046-bp PCR fragment (fragment A), covering exon 2 (255 bp), intron 2 (273 bp), and exon 3 (270 bp), and a part of introns 1 (101 bp) and 3 (100 bp). The second PCR product was 1,297-bp long (fragment B), containing complete intron 3 (596 bp), exon 4 (270 bp), intron 4 (102 bp), and exon 5 (141 bp), and a part of exons 3 (129 bp) and intron 5 (13 bp). The PCR for fragment A (25 µl) contained gDNA (100 ng), 0.8 µM of the primers (table 1), 2.5-3 mM $MgCl_2$, 0.4 mM of each dNTP, and 5 U of *Taq* polymerase. A total of 33 cycles were run, each cycle consisting of 30 s at 95°C, 50 s at 61.3°C, and 30 s at 72°C with a final extension of 8 min at 72°C. PCR for fragment B (50 µl) contained gDNA (100 ng),

0.8 µM of the primers (table 1), 2.25-2.75 mM $MgCl_2$, 0.4 mM of each dNTP and 5 U of *Taq* polymerase. A total of 28 cycles were run, each cycle consisting of 30 s at 95°C, 30 s at 58.5°C, and 40 s at 72°C with a final extension of 7 min at 72°C. The PCR products were purified, and sequenced directly on an ABI 3100 genetic analyzer (Applied Biosystems) using the PCR primers and specific internal sequencing primers (table 1). Cycle sequencing reactions were carried out with ABI Prism BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using 0.2 µM primer, 1 µl BigDye terminator, and 5x sequencing dilution buffer (400 mM Tris-HCl, 10 mM MgCl2) in a 10 µl reaction. The sequences were analyzed using the Sequence Navigator program (Applied Biosystems). At least two independent PCR reactions were performed to confirm the allele.

MIC allele characterization by T/A-cloning and sequencing

The PCR for fragment B was performed as described above with the exception that in the PCR program the final extension step is changed into 30 min at 72°C. For *MIC* allele characterization the PCR products of fragment B were purified, ligated into the plasmid vector pTZ57R/T, and utilized in the transformation using the Ins T/A clone PCR product cloning Kit (Fermentas, St. Leon-Rot, Germany). Clones were isolated by the minipreparation protocol of Qiagen and sequenced as mentioned in the section above. Sequences reported in this publication have been deposited in the EMBL database (accessionnumbers AJ748822-AJ748831).

Phylogenetic analysis

Neighbor-joining (NJ) trees were constructed with the PAUP* program version 4.0b10 for Macintosh, using the methods of Jukes-Cantor correction for multiple hits (26) for figure 1, and Kimura's two-parameter model (27) for figure 7. Bootstrap values were based on 1,000 resamplings. The PAUP* program was also used to construct phylogenetic trees using the maximum likelihood method (full heuristic search and quartet puzzling). These trees gave topologies similar to the NJ trees.

Nomenclature

The *Patr-MICA/B* alleles received official designations. The first two digits identify the lineage, while the third and fourth digits reflect the order in which the alleles were found. The fifth digit reflects a mutation in the non-coding region. Sequences will be listed in the IMGT/MHC database (28).

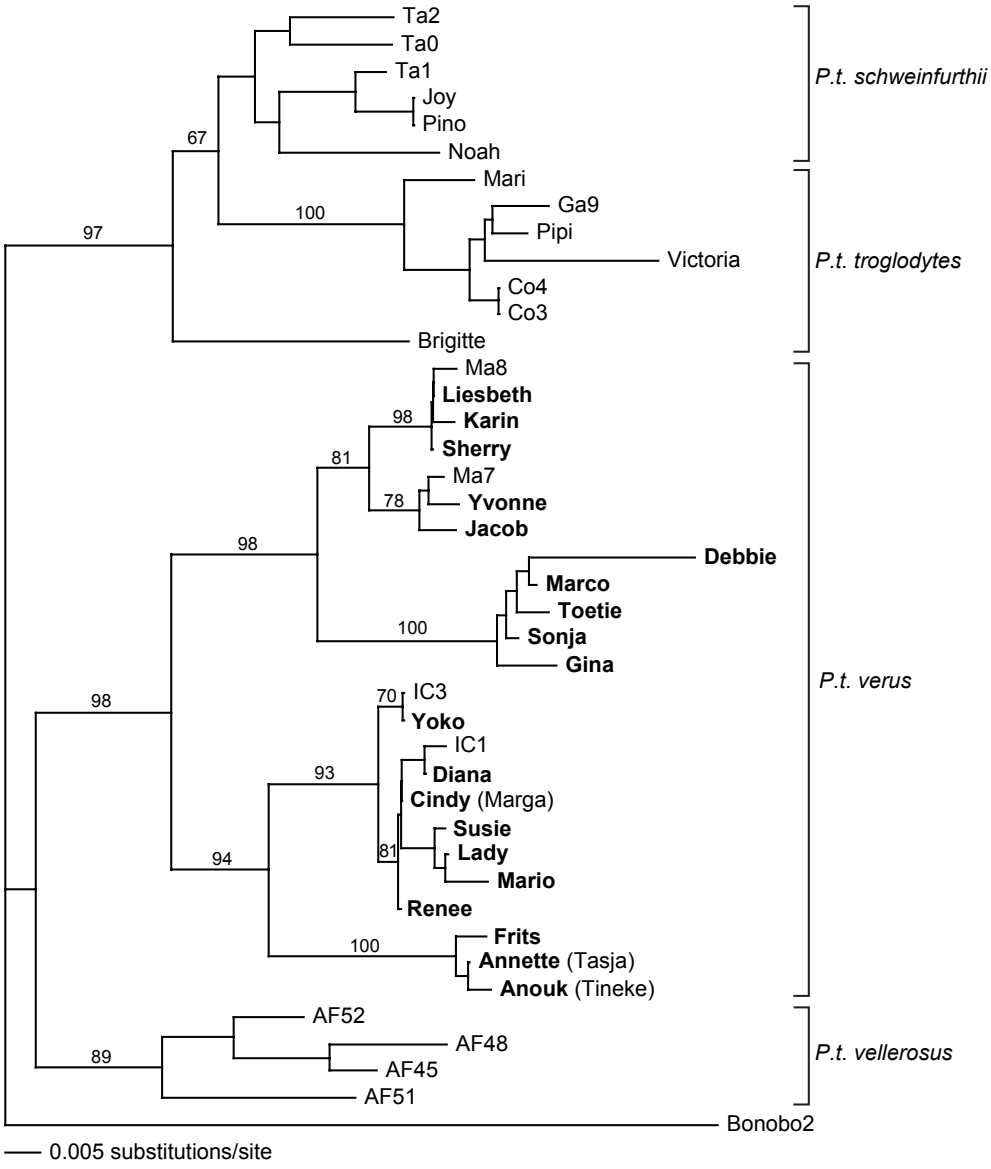


Fig. 1. Phylogenetic tree of the different mtDNA D-loop sequences detected in the BPRC founder animals (indicated in bold), and in chimpanzees of the subspecies *Pan troglodytes schweinfurthii* (Joy, Pino, and Noah) and *Pan troglodytes troglodytes* (Pipi, Victoria, and Brigitte). For comparison some sequences from two other databases were added (5, 23). From three founder animals (Marga, Tasja, and Tineke) no material was left, and in these cases the mtDNA D-loop was determined by analyzing the offspring. Brigitte is assigned to the *P.t.troglodytes* because specific nucleotides in the D-loop that are characteristic to the subspecies were found in the sequence of this animal. Relevant bootstrap values based on 1,000 resamplings are indicated.

Results and Discussion

mtDNA and MIC gene analysis in chimpanzees

From the 35 founder animals of the BPRC colony, 30 chimpanzees were selected for the present project. Phylogenetic analysis of the mtDNA D-loop sequences shows that most of the founder animals are not related (fig. 1). Although identical mtDNA D-loop sequences were detected in some chimpanzees, these animals appear to possess different MHC haplotypes (fig. 2).

Fifty-three chimpanzees (30 founder animals and 23 offspring) were screened by RSCA for *MIC* gene-associated polymorphisms. Three different fragments, covering exon 2 and 3 of the gene, were detected. The nucleotide sequence corresponding to the 2B/3B fragment is identical to the earlier published *Patr-MIC1* allele (16), which is renamed *Patr-MICA/B*01011*. The 2A/3A and 2C/3B fragments correspond with the *Patr-MICA/B*0102* and **0103* alleles, respectively. Several additional chimpanzees were analyzed by direct sequencing, resulting in the detection of three other alleles, which are designated *Patr-MICA/B*0104*, **0105*, and **0106* (figs. 3 and 4). The *Patr-MICA/B*0104* and **0106* alleles are present in the subspecies *P.t.troglodytes*, whereas *Patr-MICA/B*0105* is detected in *P.t.schweinfurthii*. All six alleles share an identical intron 2, reflecting their common ancestry.

Founder animal	mtDNA	MHC haplotype
Frits	A	a1 / b1
Carolina	B	c1 / ?
Diana	C	c2 / d2
Lady	D	? / d3
Louise	A	c4 / d4
Regina	E	c1 / d5
Sherry	F	c6 / d6
Tasja	G	c7 / d7
Marga	H	c8 / d8
Tineke	I	c1 / ?

Founder animal	mtDNA	MHC haplotype
Renza	F	c12 / ?
Sonja	J	c1 /d13
Wodka	F	c14 / d14
Yoko	K	c15 / d15
Susie	B	c18 / d18
Toetie	L	c19 / d19
Liesbeth	M	c21 / d21
Marco	N	a3 / c8
Karin	O	c23 / d23
Pearl	K	c24 / d24

Founder animal	mtDNA	MHC haplotype
Mario	P	a4 / ?
Isaac	F	a5 / b5
Debbie	Q	a3 / d26
Yvonne	R	c27 / d27
Igor	F	a6 / ?
Jacob	S	a7 / b7
Gina	E	c29 /d29
Jolanda	B	c29 /d30
Renee	T	a3 /d32
Gerrit	NT	a2 / b2

Fig 2. List of the 30 founder animals of the BPRC chimpanzee colony, showing their mtDNA typing and MHC haplotypes. The capital letters indicate the different mtDNA sequences that are detected, whereas lowercase letters depict the different MHC haplotypes. The sharing of identical mtDNA typing or MHC haplotype between different animals is indicated by colors. "?" Indicates that this part of the haplotype is not known. "NT" means not typed.

In the classical MHC class I genes most polymorphism is confined to exons 2 and 3. In humans, however, *MICA* polymorphism is not exclusively restricted to these exons, but is also located in exon 4 (29). Therefore the study was extended by sequencing intron 3 to exon 5, resulting in the definition of an additional allele, *Patr-MICA/B*0107* (fig. 3), which is present in *P.t.troglodytes*. Furthermore, intron 3 and 4 show allelic variations that appears to be caused by point mutations (fig. 4). *Patr-MICA/B*0101* is found in combination with four different intron 3 sequences, as reflected by the designations *Patr-MICA/B*01011*, **01012*, **01013*, and **01014*, respectively (fig. 4). The *Patr-MICA/B*01011* allele is observed in all three subspecies, *Patr-MICA/B*01012* is present in the subspecies *P.t.troglodytes* and *P.t.schweinfurthii,* while *Patr-MICA/B*01013* and **01014* are only detected in *P.t.verus* and *P.t.troglodytes*, respectively. Intron 3 of *Patr-MICA/B*01011* is found in combination with other alleles, whereas *Patr-MICA/B*0105* has its own unique intron 3 sequence (fig. 4). Strong conservative evolution must have operated on the *Patr-MICA/B*0101* allele. This is supported by the fact that this allele is present in all subspecies, that it is the most frequently detected allele in the population (94% of the West African chimpanzees analyzed possess at least one copy of the allele), and that it exhibits more sequence variation in the introns than in the exons.
All presently known *Patr-MICA/B* alleles share the same characteristic hybrid sequence structure. The first part (exon 2 and intron 2) is comparable to the human *MICA* sequence, whereas the second part (exon 3 to 5) is roughly equivalent to the human *MICB* sequence (fig. 3). The recombinant character of the *Patr-MICA/B* sequences is further illustrated by the absence of the GCT repeat (alanine) in exon 5. The presence of this repeat is typical for the human *MICA* alleles but as such is absent in *MICB* (fig. 3).

The MIC gene: a comparison between humans, chimpanzees, and other species

Figure 5A shows a schematic representation of the distribution of the *MICA* and *MICB* gene and their haplotypes in different species. For gorilla (*Gogo*), orangutan (*Popy*), and rhesus macaque (*Mamu*) a *MICA*-like gene has been documented (30, 31), additionally short tandem repeat analysis provided further evidence for the presence of human *MICA* gene orthologues in these species (data not shown). Recently, the genomic organization of the rhesus macaque MHC was published, showing the presence of a *MIC1* and *MIC2* gene that are situated on the chromosome on locations comparable to those of the human *MICA* and *MICB* gene (31). Furthermore, *Mamu-MIC1* possesses the characteristic GCT-repeat coding for alanine residues (fig. 3), although the alanine residues are interrupted by one valine. Thus, *MIC1* seems to be the orthologue of the human *MICA* gene. In *Mamu-MIC2* the repeat is absent, and, additionally, intron 3 shows features that are characteristic for *MICB* (fig. 6, gray boxes). Despite evidence that an ancestor of humans, great apes, and Old World monkeys possessed a *MICA*-like and *MICB*-like gene tandem on a haplotype (fig. 5A), it is evident that these genes diversified considerably in humans versus rhesus macaques. It is not known whether a *MICB*-like gene is present in gorillas or orangutans. We therefore designed primers amplifying partial *MICB* intron 3 (nucleotide position 235-513),

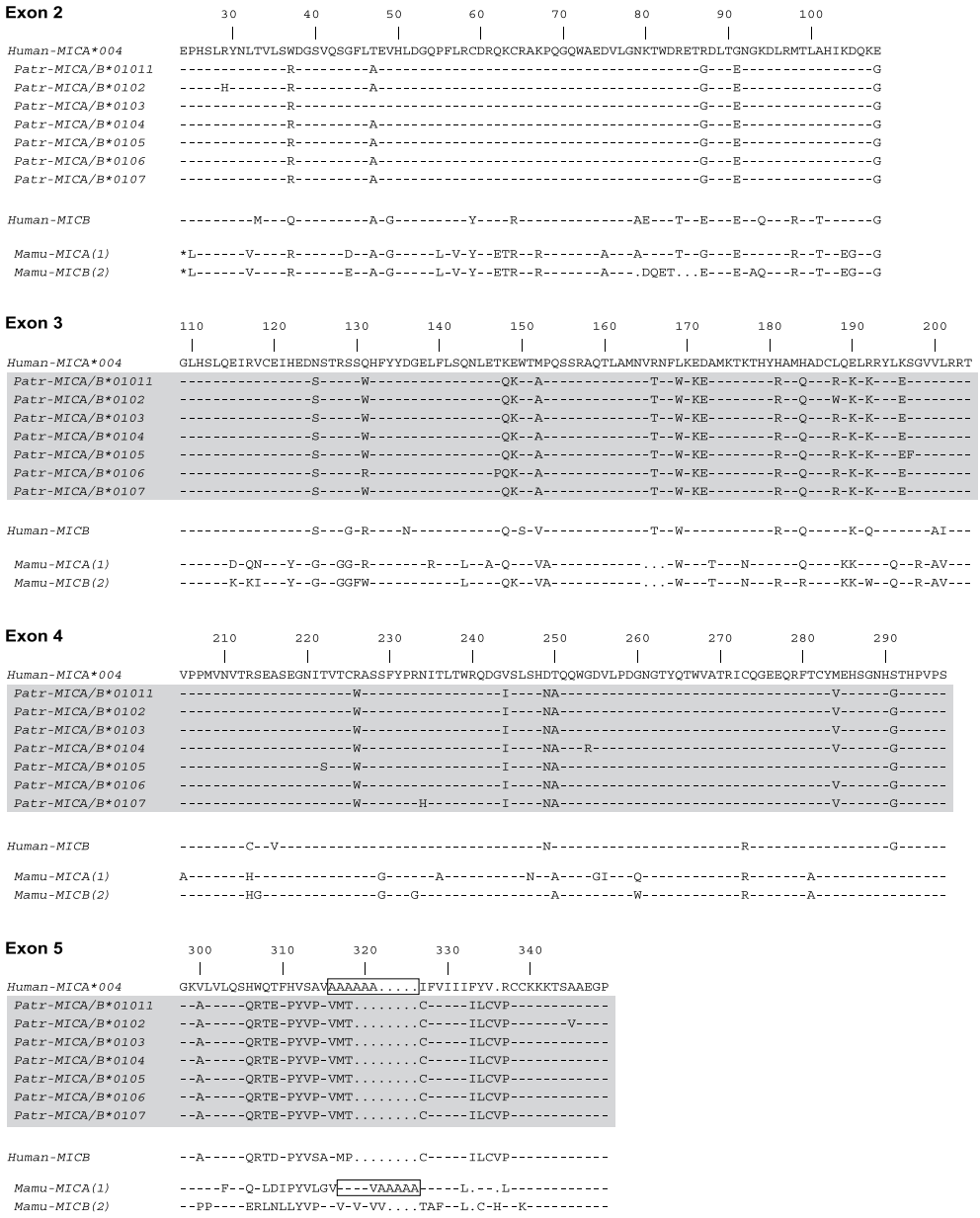


Fig. 3. Deduced amino acid sequences of the *Patr-MICA/B* alleles. Human *MICA*004* (X92841) is used as a reference, and human *MICB* (U65416) and *Mamu-MICA(1)* and *B(2)* (AC148666 and AC148679) are given for comparison. Identity to *MIC*004* is indicated by dashes, whereas amino acid replacements are depicted by the conventional one-letter code. Dots indicate a deletion. The alleles *Patr-MIC1* (AF055384) and *Patr-MIC1*02* (AF322858) were renamed *Patr-MICA/B*01011* and *Patr-MICA/B*0102*, respectively. The gray boxes indicate the part of the *Patr-MICA/B* gene that is more comparable to the human *MICB* gene, whereas exon 2 of the *Patr-MICA/B* gene is more related to human *MICA*. The open boxes in exon 5 indicate the alanine-repeat present in the human *MICA* and *Mamu-MICA(1)* gene. In humans this repeat can vary from 2 to 10 repeats per allele (28).

	Exon 2		Exon 3						Intron 3							Exon 4				Intron 4	Exon 5		
	86	139	391	438	491	561	590		36	37	81	190	343	385	663	699	759	849	85	899	1019	
Concensus	G	G	T	A	C	C	C		A	A	T	A	G	A	A	A	G	G	G	A	C	sub-species
*Patr-MICA/B*01011*	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	ptv/ptt/pts
*Patr-MICA/B*01012*								2	-	-	-	G	-	-								ptt/pts
*Patr-MICA/B*01013*								3	G	-	-	-	-	-								ptv
*Patr-MICA/B*01014*								4	-	-	-	-	A	-								ptt
*Patr-MICA/B*0102*	A	-	-	-	-	T	-		-	-	-	-	-	-	-	-	-	-	-	-	T	ptv
*Patr-MICA/B*0103*	-	A	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	A	-	-	ptv
*Patr-MICA/B*0104*	-	-	-	-	t	-	-		-	-	-	-	-	-	-	-	A	-	-	-	-	ptt
*Patr-MICA/B*0105*	-	-	-	-	-	-	T	5	-	G	C	-	-	G	T	-	-	A	-	G	-	pts
*Patr-MICA/B*0106*	-	-	C	C	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	ptt
*Patr-MICA/B*0107*	-	-	-	-	-	-	-		-	-	-	-	-	-	-	C	-	-	-	-	-	ptt

Fig. 4. Polymorphic nucleotide positions in different exons and introns of the *Patr-MICA/B* alleles. Intron 2 is identical for all the detected alleles. Identity to the consensus sequence (depicted at the top) is indicated by dashes. In the exons, a capital letter indicates a nonsynonymous substitution, whereas a lowercase letter indicates a synonymous substitution. The numbers above the consensus indicate the nucleotide substitution position. The introns are numbered individually, whereas the exons are numbered through.

including the characteristic insert (fig. 6). In orangutans we were able to amplify this sequence, indicating that at least a *MICB*-like gene segment is present in this species (fig. 5A). However, the primers did not produce a product with gorilla samples. This could indicate that a *MICB*-like gene is absent in gorillas, although the possibility that the primers are not working optimally cannot be excluded.

Exon 2 of the human *MICA* and *B* and the *Patr-MICA/B* nucleotide sequences were subjected to phylogenetic analysis together with the known gorilla, orangutan, and rhesus macaque *MICA* and *MICB* orthologues (fig. 7). The tree illustrates that this part of the *Patr-MICA/B* gene is distantly related to the human *MICA* gene, but shows a reduced level of variation of one lineage versus four lineages, respectively. The second part of the *Patr-MICA/B* sequences, exon 3 to 5, showed a more intermediate character versus human *MICA* and *MICB* (data not shown). Comparison of the data shows that intron 3 of the *Patr-MICA/B* alleles possesses features more characteristic of the human *MICB* gene (fig. 6), and in addition exon 5 illustrates the more close genetic relation of the chimpanzee and human *MICB* sequences (fig. 3). The reported chimpanzee class I region is characterized by a 95-kb genomic deletion resulting in a *Patr-MICA/B* fusion product (fig. 5B) (18, 21). The present communication shows that all chimpanzees analyzed so far possess haplotypes with the hybrid *MICA/B* gene. *MICA* polymorphism studies in humans resulted in the identification of 54 different alleles (12, 33-36), whereas for *MICB* 14 different alleles are detected (37-40). In the 30 unrelated West African chimpanzees studied, only four *MICA/B* alleles were identified (fig. 4), of which only three encode slightly distinct proteins (fig. 3). All variation can be explained by point mutations. Likewise, in the two other chimpanzee subspecies studied, only hybrid *MICA/B* sequences are detected. Taken together, 10 *Patr-MICA/B* alleles were identified, among which 7 encode for distinct proteins. Thus, it seems that West African chimpanzees have only one functional copy of a *MIC* gene, which possesses only one lineage and limited allelic variation, and the same trend seems to be observed in chimpanzees of other subspecies.

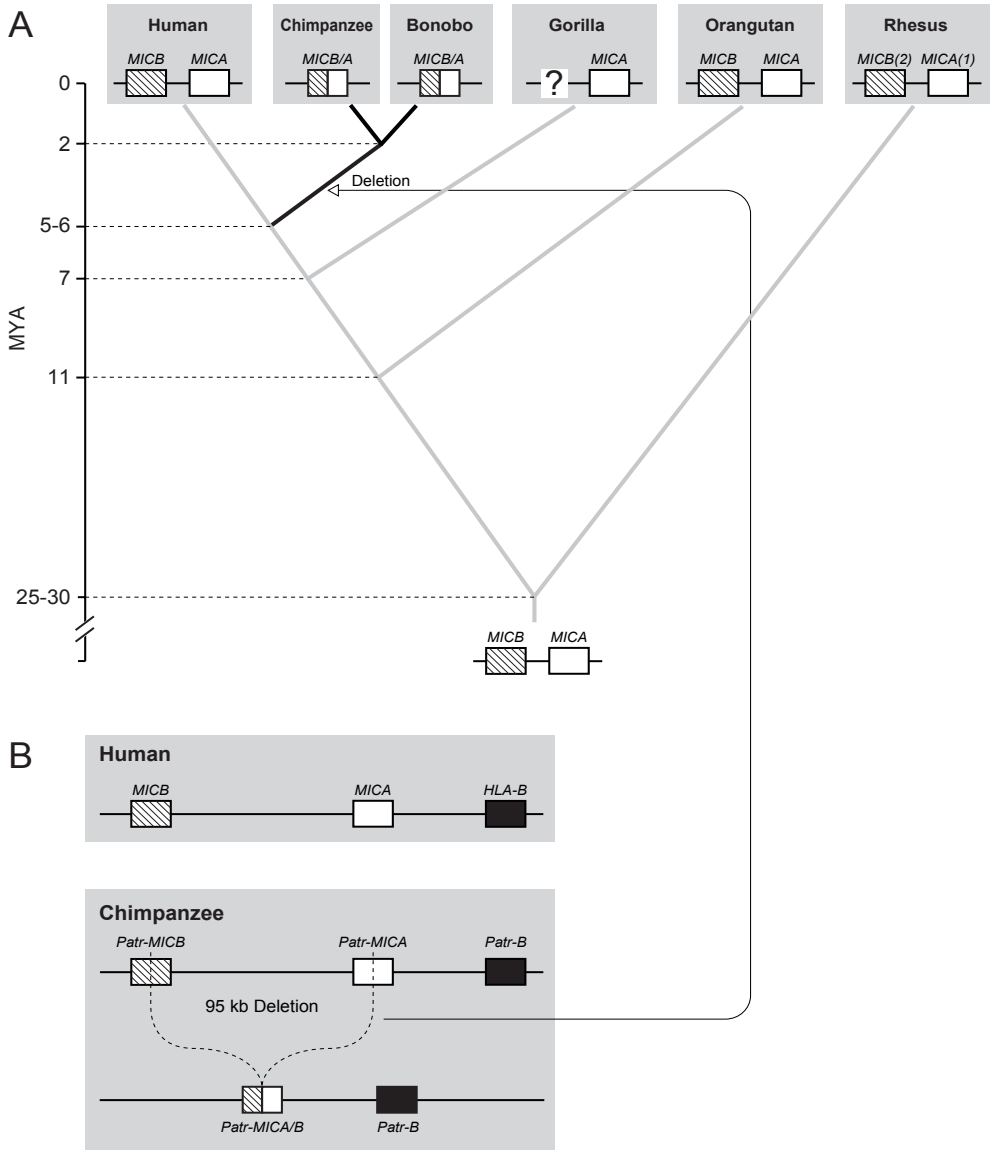
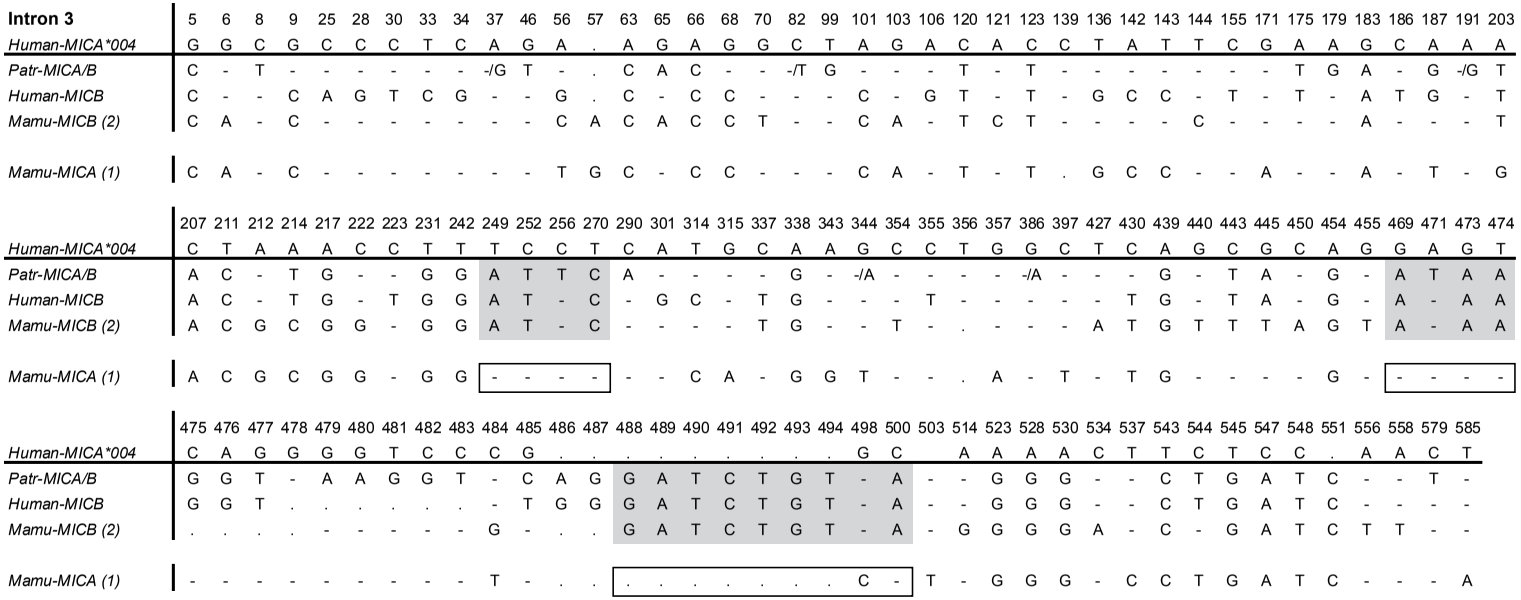


Fig. 5. (A) Schematic representation of the relationship of the *MICA* and *MICB* gene in different species. The "?" indicates that it is not known if a *MICB*-like gene is present in this species. (B) A 95kb genomic deletion in the β block of the chimpanzee producing a *Patr-MICA/B* fusion product, the schematic representation is based on previously published work (18), and compared to the human organization. MYA, million years ago.

Analysis of the *Patr-MICA/B* gene introns 3 and 4 revealed limited allelic variation, five and two different sequences, respectively. In the 595 nucleotide bases of intron 3, only one (*Patr-MICA/B*01012* and *Patr-MICA/B*01014*) or three (*Patr-MICA/B*0105*) nucleotide substitutions are observed (fig. 4). The existence of these few nucleotide substitutions

indicates that the alleles are probably of relatively recent origin.
One question to be answered is: when did the 95-kb deletion resulting in a *Patr-MICA/B* fusion product take place? Most likely this deletion occurred after the human/chimpanzee split (fig. 5A). The presence of an equivalent in bonobos indicates that the deletion happened before the speciation of chimpanzee and bonobo approximately 2 MYA. Thus, the deletion took place in the interval of 2 to 5-6 MYA (fig. 5A). The alternative explanation, that the human genome experienced a 95-kb insertion generating the *MICA* and *MICB* genes, is highly unlikely. Such an assumption is also in disagreement with the fact that the existence of the *MICA* gene predates the speciation of humans, chimpanzees, gorillas, orangutans (30), and rhesus macaques (fig. 5A) (31). This would imply that all ancestral chimpanzee haplotypes with the original *Patr-MICA* and *Patr-MICB* genes have been subject to a strong negative selection because these genes are apparently absent in the contemporary population. Earlier we reported that both the MHC class I *Patr-A* and *Patr-B* loci have suffered a severe repertoire reduction with regard to lineages, as compared to the human population. The intron 2 analysis suggested that the selective sweep in the MHC region occurred approximately 2 to 3 MYA (11). Due to the close vicinity of the *Patr-B* and *Patr-MICA/B* loci it is possible that the hybrid *Patr-MICA/B* gene was selected based on a piggyback effect.
In humans, four candidate genes are found within the 95-kb section that is deleted: namely two uncharacterized transcripts, 3.8-1.1 and P5-1 (class I-like transcript), and two pseudogenes, HCGIX-1 and HLA-X (18, 21). Based on the knowledge that humans and chimpanzees shared a common ancestor, it is highly probable that equivalents of these genes were also present in chimpanzees before the deletion. However, since the function of the 3.8-1.1 and P5-1 transcripts is unknown, we cannot exclude the possibility that the deletion of these genes may have given a selective advantage to chimpanzee haplotypes with the hybrid *Patr-MICA/B* gene.

Linkage between the MHC-B locus and the human MICA/Patr-MICA/B locus
In humans, particular *MICA* alleles are strongly associated to particular *HLA-B* alleles. From an evolutionary point of view one could expect that particular *Patr-MICA/B* alleles also show an association with certain *Patr-B* alleles. Genomic mapping studies demonstrated that the *Patr-MICA/B* gene and the human *MICA* gene are located 45.2 and 46.4 kb centromeric from the *MHC-B* locus, respectively. However, in the well-characterized BPRC chimpanzee population, specific association between particular *Patr-MICA/B* alleles and particular *Patr-B* alleles is not observed. The most frequently detected *MIC* allele, *Patr-MICA/B*01011*, is present in all subspecies. Furthermore, all the *Patr-MICA/B* alleles that are detected seem to belong to one lineage. Again, it is possible that the *Patr-B* lineages/alleles that survived the ancient selective sweep were all linked to the frequently detected *Patr-MICA/B*01011* allele and that the contemporary limited allelic variation seen in the *Patr-MICA/B* locus was generated after the selective sweep. Support for this observation is found in the *Patr-MICA/B* intron 3 sequences (fig. 4), where only a few nucleotide substitutions distinguish the different alleles. This most probably reflects their more recent origin.



Intron 3	5	6	8	9	25	28	30	33	34	37	46	56	57	63	65	66	68	70	82	99	101	103	106	120	121	123	139	136	142	143	144	155	171	175	179	183	186	187	191	203
*Human-MICA*004*	G	G	C	G	C	C	C	T	C	A	G	A	.	A	G	A	G	G	C	T	A	G	A	C	A	C	C	T	A	T	T	C	G	A	A	G	C	A	A	A
Patr-MICA/B	C	-	T	-	-	-	-	-	-	-/G	T	-	.	C	A	C	-	-	-/T	G	-	-	-	T	-	T	-	-	-	-	-	-	-	T	G	A	-	G	-/G	T
Human-MICB	C	-	-	C	A	G	T	C	G	-	-	G	.	C	-	C	C	-	-	-	C	-	G	T	-	T	-	G	C	C	-	T	-	T	-	A	T	G	-	T
Mamu-MICB (2)	C	A	-	C	-	-	-	-	-	-	-	C	A	C	A	C	C	T	-	-	C	A	-	T	C	T	-	-	-	-	C	-	-	-	-	A	-	-	-	T
Mamu-MICA (1)	C	A	-	C	-	-	-	-	-	-	-	T	G	C	-	C	C	-	-	-	C	A	-	T	-	T	.	G	C	C	-	-	A	-	-	A	-	T	-	G

	207	211	212	214	217	222	223	231	242	249	252	256	270	290	301	314	315	337	338	343	344	354	355	356	357	386	397	427	430	439	440	443	445	450	454	455	469	471	473	474
*Human-MICA*004*	C	T	A	A	A	C	C	T	T	T	C	C	T	C	A	T	G	C	A	A	G	C	C	T	G	G	C	T	C	A	G	C	G	C	A	G	G	A	G	T
Patr-MICA/B	A	C	-	T	G	-	-	G	G	A	T	T	C	A	-	-	-	-	G	-	-/A	-	-	-	-	-/A	-	-	-	G	-	T	A	-	G	-	A	T	A	A
Human-MICB	A	C	-	T	G	-	T	G	G	A	T	-	C	-	G	C	-	T	G	-	-	-	T	-	-	-	-	-	T	G	-	T	A	-	G	-	A	-	A	A
Mamu-MICB (2)	A	C	G	C	G	G	-	G	G	A	T	-	C	-	-	-	-	T	G	-	-	T	-	.	-	-	-	A	T	G	T	T	T	A	G	T	A	-	A	A
Mamu-MICA (1)	A	C	G	C	G	G	-	G	G	-	-	-	-	-	-	C	A	-	G	G	T	-	-	.	A	-	T	-	T	G	-	-	-	-	G	-	-	-	-	-

	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	498	500	503	514	523	528	530	534	537	543	544	545	547	548	551	556	558	579	585
*Human-MICA*004*	C	A	G	G	G	G	T	C	C	C	G	.	.	.	.	.	.	.	.	.	G	C		A	A	A	A	C	T	T	C	T	C	C	.	A	A	C	T
Patr-MICA/B	G	G	T	-	A	A	G	G	T	-	C	A	G	G	A	T	C	T	G	T	-	A	-	-	G	G	G	-	-	C	T	G	A	T	C	-	-	T	-
Human-MICB	G	G	T	.	.	.	.	.	.	-	T	G	G	G	A	T	C	T	G	T	-	A	-	-	G	G	G	-	-	C	T	G	A	T	C	-	-	-	-
Mamu-MICB (2)	.	.	.	.	-	-	-	-	-	G	-	.	.	G	A	T	C	T	G	T	-	A	-	G	G	G	G	A	-	C	-	G	A	T	C	T	T	-	-
Mamu-MICA (1)	-	-	-	-	-	-	-	-	-	T	-	.	.	.	.	.	.	.	.	.	C	-	T	-	G	G	G	-	C	C	T	G	A	T	C	-	-	-	A

Fig. 6. Polymorphic nucleotide positions in intron 3. The *Patr-MICA/B*, human *MICB*, and *Mamu-MICA(1)* and *B(2)* are compared to the human *MICA*004*. Identity is indicated by dashes. Dots indicate deletions. For instance, "-/A" means that differences in the reported alleles are detected. The numbers above the human *MICA*004* alignment indicate the nucleotide substitution positions. The open and gray boxes indicate the parts that show the differentiation between *MICA* and *B*, respectively.

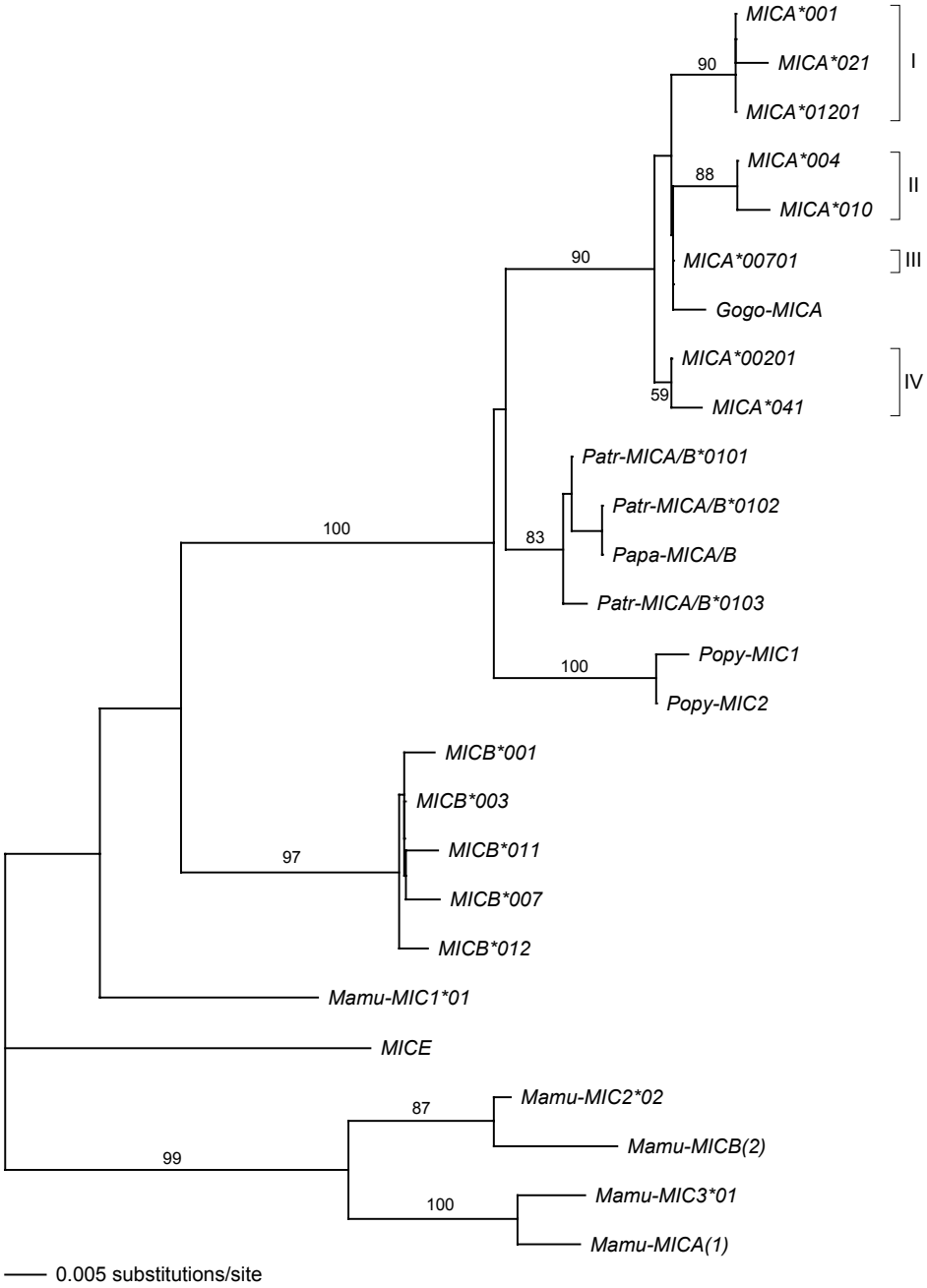


Fig. 7. Phylogenetic tree (exon 2) of the *Patr-MICA/B* alleles and human *MICA* and *MICB* alleles. Added are gorilla (AF045597), orangutan (AF045599 and AF045600), and rhesus macaque (AJ242438, AJ242440, AJ242441, AC148666, and AC148679) *MICA*-like and *MICB*-like sequences. The human *MICE* (AP000521) allele is taken as an out-group. The brackets indicate the division of the human *MICA* alleles into different lineages. The tree shows that the *Mamu-MIC1*01* (32) and the *Mamu-MICA(1)* (31) alleles diversified. Relevant bootstrap values based on 1,000 resamplings are indicated.

Could Patr-MICA/B contribute to AIDS resistance in chimpanzees?

Chimpanzees can be naturally infected with Simian immunodeficiency virus from chimpanzees (SIV_{cpz}) but are also susceptible to infection with Human immunodeficiency virus type 1 (HIV-1). Although natural SIV_{cpz} infections in West African chimpanzees have never been observed, the animals are susceptible to HIV-1 infection in captivity (41). In contrast to humans, chimpanzees normally do not develop symptoms of acquired immunodeficiency syndrome (AIDS). We have put forward the hypothesis that in the past chimpanzees may have been decimated by an AIDS-like pandemic caused by an HIV-1/SIV_{cpz}-like retrovirus and that the contemporary population represents the offspring of the survivors (11). In this respect, the contemporary chimpanzee population may have been enriched for resistance genes, and a candidate group that was identified is the group of MHC molecules, which play an important role in controlling infections.

Chimpanzees are often described as herbivores, but they also hunt and consume other nonhuman primate species (42). As most of these Old World monkeys have natural SIV infections, chimpanzees may have become infected through predating on infected monkeys. Probably chimpanzees have been infected multiple times and by various SIV-like strains (43, 44). To date little is known about the transmission of SIV_{cpz} between different chimpanzees. The initial infections with HIV-1/SIV_{cpz}-like viruses in humans probably happened through blood contact during the hunting of infected chimpanzees. The routes of HIV-1 infection in humans are now known to be via intimate sexual contact, contaminated blood or blood products, or by transmission from mother to child (45). Nonetheless, the intestine may be the major site for HIV replication and depletion of $CD4^+$ T cells (46). In this perspective, it is possible that the hybrid *Patr-MICA/B* gene is highly efficient in detecting stress mediated by infection, and as such may induce efficient natural killer (NK) cell-mediated killing.

The human *MICA* and probably also the closely related *MICB* molecules are ligands for the NKG2D receptor, which is expressed on NK cells, $\gamma\delta$ T cells, and $CD8^+$ $\alpha\beta$ T cells. After cellular stress, triggered for instance by a viral infection or malignant transformation, *MIC* is upregulated and may provoke an immune response (14, 15). Recent work, for instance, showed that the activating KIR gene KIR3DS1 is associated with a delay in progression to AIDS in HIV-1-infected individuals (47). The chimpanzee MICA/B molecule is able to recognize human Vδ1 $\gamma\delta$ T cells specific for MICA and MICB, suggesting a conserved recognition site (16). Indeed, also the NKG2D receptor shows a high degree of similarity, approximately 98.9%, between humans and chimpanzees (48). Based upon these observations and the knowledge that humans and chimpanzees are closely related, a similar kind of immune response as described in humans is plausible.

The possible role of MIC-elicited anti-HIV-1/SIV_{cpz} NK effector responses remains to be proved, and subsequent studies will be required to elucidate the functional significance of the *Patr-MICA/B* gene in virus infections.

Acknowledgements

The authors wish to thank D. Devine and H. van Westbroek for expert assistance, J. Heeney for discussion, and S. Langehuizen for technical assistance. This study was financed in part by the European Union Project, IMGT-QLG2-2000-01287.

Literature cited

1. Bahram S, Bresnahan M, Geraghty DE, Spies T (1994) A second lineage of mammalian major histocompatibility complex class I genes. Proc Natl Acad Sci U S A 91:6259-6263.
2. Leelayuwat C, Townend DC, Degli-Esposti MA, et al. (1994) A new polymorphic and multicopy MHC gene family related to nonmammalian class I. Immunogenetics 40:339-351.
3. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51:410-424.
4. de Groot NG, Otting N, Arguello R, et al. (2000) Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. Immunogenetics 51:398-409.
5. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
6. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713.
7. Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.
8. Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A 97:11354-11358.
9. Yu N, Jensen-Seaman MI, Chemnick L, et al. (2003) Low nucleotide diversity in chimpanzees and bonobos. Genetics 164:1511-1518.
10. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
11. de Groot NG, Otting N, Doxiadis GG, et al. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci U S A 99: 11748-11753.
12. Bahram S (2000) MIC genes: from genetics to biology. Adv Immunol 76:1-60.
13. Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279:1737-1740.
14. Bauer S, Groh V, Wu J, et al. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285:727-729.
15. Gleimer M, Parham P (2003) Stress management: MHC class I and class I-like molecules as reporters of cellular stress. Immunity 19:469-477.
16. Steinle A, Groh V, Spies T (1998) Diversification, expression, and gamma delta T cell recognition of evolutionarily distant members of the MIC family of major histocompatibility complex class I-related molecules. Proc Natl Acad Sci U S A 95:12510-12515.

17. Cattley SK, Longman N, Dawkins RL, et al. (1999) Phylogenetic analysis of primate MIC (PERB11) sequences suggests that the representation of the gene family differs in different primates: comparison of MIC (PERB11) and C4. Eur J Immunogenet 26:233-238.
18. Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122.
19. Radosavljevic M, Bahram S (2003) In vivo immunogenetics: from MIC to RAET1 loci. Immunogenetics 55:1-9.
20. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
21. Anzai T, Shiina T, Kimura N, et al. (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc Natl Acad Sci U S A 100:7708-7713.
22. Kocher TD, Thomas WK, Meyer A, et al. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196-6200.
23. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
24. Mendoza-Rincon J, Arguello JR, Perez-Rodriguez M, et al. (1999) Characterization of the MICA polymorphism by sequence-specific oligonucleotide probing. Immunogenetics 49: 471-478.
25. Arguello JR, Little AM, Pay AL, et al. (1998) Mutation detection and typing of polymorphic loci through double-strand conformation analysis. Nat Genet 18:192-194.
26. Jukes TH, Cantor CR (1969) Mammalian Protein Metabolism (Academic press, New York).
27. Swofford DL (1992) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods).
28. Robinson J, Waller MJ, Parham P, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31: 311-314.
29. Fodil N, Laloux L, Wanner V, et al. (1996) Allelic repertoire of the human MHC class I MICA gene. Immunogenetics 44:351-357.
30. Pellet P, Vaneensberghe C, Debre P, et al. (1999) MIC genes in non-human primates. Eur J Immunogenet 26:239-241.
31. Daza-Vamenta R, Glusman G, Rowen L, et al. (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res 14:1501-1515.
32. Seo JW, Bontrop R, Walter L, Gunther E (1999) Major histocompatibility complex-linked MIC genes in rhesus macaques and other primates. Immunogenetics 50:358-362.
33. Petersdorf EW, Shuler KB, Longton GM, et al. (1999) Population study of allelic diversity in the human MHC class I-related MIC-A gene. Immunogenetics 49:605-612.
34. Visser CJ, Tilanus MG, Tatari Z, et al. (1999) Sequencing-based typing of MICA reveals 33 alleles: a study on linkage with classical HLA genes. Immunogenetics 49:561-566.
35. Yao Z, Volgger A, Keller E, et al. (1999) Allelic variation in the intron 2 and 3 of the MICA gene. J Biol Regul Homeost Agents 13:47-50.
36. Robinson J, Perez-Rodriguez M, Waller MJ, et al. (2001) MICA sequences 2000. Immunogenetics 53:150-169.
37. Bahram S, Spies T (1996) Nucleotide sequence of a human MHC class I MICB cDNA. Immunogenetics 43:230-233.

38. Ando H, Mizuki N, Ota M, et al. (1997) Allelic variants of the human MHC class I chain-related B gene (MICB). Immunogenetics 46:499-508.
39. Pellet P, Renaud M, Fodil N, et al. (1997) Allelic repertoire of the human MICB gene. Immunogenetics 46:434-436.
40. Visser CJ, Tilanus MG, Schaeffer V, et al. (1998) Sequencing-based typing reveals six novel MHC class I chain-related gene B (MICB) alleles. Tissue Antigens 51:649-652.
41. Rutjens E, Balla-Jhagjhoorsingh S, Verschoor E, et al. (2003) Lentivirus infections and mechanisms of disease resistance in chimpanzees. Front Biosci 8:d1134-1145.
42. Stanford CB, Wallis J, Matama H, Goodall J (1994) Patterns of predation by chimpanzees on red colobus monkeys in Gombe National Park, 1982-1991. Am J Phys Anthropol 94:213-228.
43. Courgnaud V, Salemi M, Pourrut X, et al. (2002) Characterization of a novel simian immuno deficiency virus with a vpu gene from greater spot-nosed monkeys (Cercopithecus nictitans) provides new insights into simian/human immunodeficiency virus phylogeny. J Virol 76: 8298-8309.
44. Bailes E, Gao F, Bibollet-Ruche F, et al. (2003) Hybrid origin of SIV in chimpanzees. Science 300:1713.
45. Levy JA, Scott I, Mackewicz C (2003) Protection from HIV/AIDS: the importance of innate immunity. Clin Immunol 108:167-174.
46. Veazey RS, Lackner AA (2004) Getting to the guts of HIV pathogenesis. J Exp Med 200: 697-700.
47. Martin MP, Gao X, Lee JH, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet 31:429-434.
48. Shum BP, Flodin LR, Muir DG, et al. (2002) Conservation and variation in human and common chimpanzee CD94 and NKG2 genes. J Immunol 168:240-252.

Chapter 5

Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics

Natasja G. de Groot[1], Corrine M.C. Heymans[1], Nanine de Groot[1], Nel Otting[1], Annemiek J.M. de Vos-Rouweler[1], Edmond J. Remarque[1], Maxime Bonhomme[2], Gaby G.M. Doxiadis[1], Brigitte Crouau-Roy[2], Ronald E. Bontrop[1]

Molecular Ecology 17 (2008), 2074-2088

[1]Department of Comparative Genetics and Refinement, and Parasitology, Biomedical Primate Research Centre, Rijswijk, The Netherlands.
[2]Laboratoire Evolution et Diversité Biologique (EDB), Université Paul Sabatier, Toulouse, France.

Abstract

Chimpanzees experienced a reduction of the allelic repertoire at the major histocompatibility complex (MHC) class I *A* and *B* loci, which may have been caused by a retrovirus belonging to the simian immunodeficiency virus (SIV) family. Extended MHC haplotypes were defined in a pedigreed chimpanzee colony. Comparison of genetic variation at microsatellite markers mapping inside and outside the *Mhc* region was carried out in humans and chimpanzees to investigate the genomic extent of the repertoire reduction. Multilocus demographic analyses underscore that chimpanzees indeed experienced a selective sweep that mainly targeted the chromosomal segment carrying the *Mhc* class I region. Probably due to genetic linkage, the sweep also affected other polymorphic loci, mapping in the close vicinity of the *Mhc* class I region genes. Nevertheless, although the allelic repertoire at particular *Mhc* class I and II loci appears to be limited, naturally occurring recombination events allowed the establishment of haplotype diversity after the sweep. However, recombination did not have sufficient time to erase the signal of the selective sweep.

Introduction

The major histocompatibility complex (MHC), designated in humans as human leukocyte antigen (HLA), is a multigene family located on the short arm of Chromosome 6; it encodes, among others, two clusters of cell-surface proteins. MHC class I gene products play a role in the defence against intracellular pathogens (1), whereas class II allotypes regulate antibody responses directed against pathogens of extracellular origin (2). The hallmark of MHC gene products is their extensive degree of polymorphism, mostly confined to the amino acid residues that map to the peptide binding site (3). In humans, the classical class I genes are designated *HLA-A*, *-B* and *-C*, whereas *Patr-A*, *-B* and *-C* represent the orthologues in chimpanzees. The class II region genes are referred to as *HLA-DP*, *-DQ* and *-DR* and *Patr-DP*, *-DQ* and *-DR*, respectively. Chimpanzees and humans are closely related, displaying approximately 98.7% similarity at the nonrepetitive DNA level, and both species shared an ancestor approximately 5-6 million years ago (Ma) (4). Despite the high degree of similarity and the fact that many *Mhc* lineages predate speciation processes, most of the alleles themselves are of more recent origin and are unique to a given species (5-7).

Three different subspecies of chimpanzees are officially recognised. In two subspecies, *Pan troglodytes troglodytes* (*P.t.t.*) and *P.t. schweinfurthii* (*P.t.s.*), natural infections with chimpanzee-derived simian immunodeficiency virus (SIV_{cpz}), the closest relative of human immunodeficiency virus type 1 (HIV-1), have been established (8, 9). Next to humans, chimpanzees are the only species susceptible to infection with HIV-1. However, like a small subset of human individuals, chimpanzees in general do not develop symptoms of acquired immunodeficiency syndrome (AIDS) after natural or experimental infection

with HIV-1 or SIV_{cpz} strains (10, 11). Although natural infections with SIV_{cpz} in the *P.t. verus* (*P.t.v.*) subspecies have not been documented, some animals belonging to this subspecies do mount effective cytotoxic T cell (CTL) responses to conserved HIV-1 epitopes in the context of particular MHC class I molecules (12).

Earlier studies have highlighted that, in comparison to humans, chimpanzees have a reduced *Mhc* class I *A* repertoire (13-15). As MHC gene products are known to experience diversifying selection, reflected by their extensive polymorphism (16, 17), it is difficult to draw firm conclusions on particular selection processes based on an analysis of isolated exon sequences. Subsequent comparative screening of intron variation in the *Mhc* class I loci has revealed that the allelic repertoire at the *B* locus is reduced profoundly. The combination of the previously described observations accumulated in the hypothesis that the reduced *Mhc* class I repertoire in chimpanzees may be due to an ancient selective sweep caused by an AIDS-like pandemic (18).

Microsatellites are valuable markers for evolution and population studies (19, 20) and can be used accordingly to determine whether a genetic region has been affected by a repertoire reduction: for instance, due to a selective sweep (21). In the present report, microsatellite markers mapping inside and outside the MHC region were analysed to define more specifically the segment and also the extent of the region that has been targeted by the sweep. The panel comprised chimpanzees of a pedigreed West-African population as well as animals originating from other colonies and/or from other subspecies and individuals selected from a Caucasoid human population. The genetic variation of the MHC markers was compared to the unlinked autosomal markers located outside the MHC region to evaluate a potential significant difference in the MHC genetic variation between chimpanzees and humans. Subsequently, multilocus demographic analyses for the MHC and non-MHC microsatellite markers within the chimpanzee sample were performed to test for a reduction in genetic variation in the MHC, which would be due to selection and/or demographic history. In addition, the impact of recombination on genetic variation over the entire chimpanzee MHC is discussed by studying the extended MHC haplotypes and analysing them in the context of the microsatellite data. The present communication provides data of interest for investigators who perform research on the host-pathogen coevolution as well as for those who are scholars of population and conservation genetics of chimpanzees.

Materials and Methods

DNA samples

The West African chimpanzee (*P.t.v*) colony studied started with 35 founder animals originating from Sierra Leone (Fig. 1). The colony increased to more than 200 animals, covering three generations, and most animals were characterised at the molecular level for *Mhc* class I and II polymorphisms (14, 22). The offspring of the founder animals were

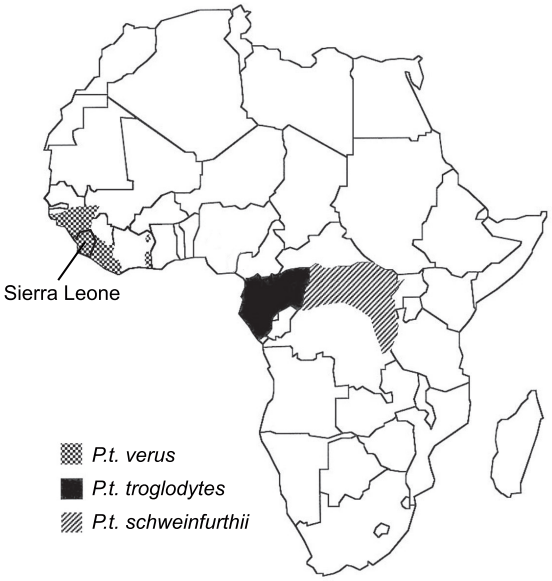


Fig. 1. Map of the African continent indicating the habitats of the different chimpanzee subspecies. There may be a fourth subspecies, Pan troglodytes vellerosus, which has its habitat in Nigeria and parts of Cameroon (66).

pedigreed based on the segregation of serological specificities (Patr-A and -B) and molecular-defined class II gene polymorphisms. For the present study, 25 haplotyped founder animals were selected. In addition, 10 individuals of the *P.t.v.* subspecies, belonging to other colonies, were studied, as well as six *P.t.t.* and five *P.t.s.* animals.

The 47 human DNA samples originating from unrelated Caucasoid individuals were provided by the department of Immunohaematology and Blood Bank of the Leiden University Medical Centre. For comparison, 25 individuals were selected at random from this cohort.

Mitochondrial DNA analysis (mtDNA)

The amplification and sequencing of the mitochondrial DNA (mtDNA) D-loop (380 bp) of the selected chimpanzees has been reported previously (23). The same protocol was used to amplify and sequence the mtDNA D-loop for the human samples, with the exception that the polymerase chain reaction (PCR) products were prepared with the ABI Prism BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems) using 0.2 µM primer, 1 µL BigDye Terminatormix and 5x sequencing dilution buffer (400mM Tris-HCl, 10 mM $MgCl_2$) in a 10 µL reaction and sequenced directly on an ABI 3100 automatic sequencer (Applied Biosystems).

Microsatellite typing of the MHC region

The microsatellite markers used for genotyping the MHC region (Table S1, Supplementary material) are documented (24), and have been tested, in part, in rhesus macaques (25).

The presence of the microsatellite markers in chimpanzees was verified using the NCBI database (www.ncbi.nlm.nih.gov/sts), and on the basis of homology (electronic-pcr) it is likely that the same microsatellite can be amplified in humans and chimpanzees with the same primerset (Fig. S1, Supplementary material). Marker D6S2811 was amplified using the protocol developed by Dr M. Carrington (www.ihwg.org/shared/micros). The other markers were amplified using a denaturation step of 5 min at 94 °C, 5 cycles of 1 min at 94 °C, 45 s at 58 °C and 45 s at 72 °C, followed by 25 cycles of 45 s at 94 °C, 30 s at 58 °C and 45 s at 72 °C, with a final extension of 30 min at 72 °C. For the markers D6S2792 and D6S2810, the annealing temperature was set to 62 °C and 63.5 °C, respectively. For D6S2972, in the case of the human samples, the annealing temperature was 52 °C. The markers D6S276, D6S291 and D6S1691 were used in a multiplex PCR mixture, using a 10-fold concentrated primer mixture. The PCR mixtures were in a total volume of 25 µL containing 50 ng of DNA, 2.5 mM $MgCl_2$ (for D6S2811 2.0 mM $MgCl_2$ is used), 0.2 mM dNTP's, 2.5 µL of 10x PCR buffer (Invitrogen) and 1 U *Taq* DNA polymerase (Invitrogen). Two microlitres of PCR product was mixed with 15 µL Hi-di Formamide and 0.5 µL GeneScan-350 ROX standard (Applied Biosystems) and the samples were run on an ABI 3130*xl* genetic analyser (Applied Biosystems). The results were analysed with Genemapper (Applied Biosystems).

Microsatellite typing of unlinked autosomal loci located outside the MHC region

The markers used for genotyping the regions located outside the MHC (Table S1, Supplementary material) are reliable for paternity testing (MC Penedo, personal communication). The NCBI database was used to verify whether these markers share homologous amplification sites in humans and chimpanzees (Fig. S1, Supplementary material). The markers were assembled in multiplexes, and the amplification took place by means of the PCR protocol described above. PCR mixtures were in a total volume of 25 µL containing 50 ng of DNA, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 2.5 µL of 10x PCR buffer, 5 µL stock solution of the multiplex primer mixture and 1 U *Taq* DNA polymerase. The sample preparation and its analysis were as described above.

Definition of MHC haplotypes

Haplotypes have been defined by studying segregation. To identify chimpanzee MHC haplotypes, direct sequencing of the *Patr-DPB1*, *DQA1* and *DQB1* loci was performed. The PCR mixtures (50 µL) contained 250 ng of DNA, 0.4 mM each dNTP, 0.8 mM of the locus-specific primers (Table S2, Supplementary material) and 5 U of *Taq* DNA polymerase. The cycling parameters consisted of an initial denaturation step of 2 min at 94 °C, 3 cycles of 90 s for each step at 94 °C, specific annealing temperature (Table S2) and 74 °C, followed by 32 cycles of 30 s for each step at the same temperatures as described above, with a final extension of 7 min at 72 °C. PCR products were purified as described (23) and sequenced with the M13 primer and the locus-specific reverse PCR primer (Table S2). Sequence samples were prepared as described in the mtDNA analysis section.

DGGE was used to separate the different *Patr-DRB* alleles, and the visualized bands were excised from the gel and subsequently sequenced using the primers 5'DRBseq and 3'DRBseq (26). The techniques used were performed as published (27). Characterisation of the *Patr-A*, *-B* and *-C* alleles has been reported in detail (14).

Statistics, microsatellite analysis programs and phylogenetic analysis

The number of unique alleles (n_e) for each microsatellite marker was calculated for humans and for chimpanzees using bootstrapping methods. Fifty haplotypes for the MHC markers, and 42 haplotypes for the non-MHC markers were re-sampled 10 000 times and were put back after every sampling, for both species separately. The difference in the number of unique alleles (Δn_e) was calculated by subtracting the n_e of the chimpanzee from the n_e of the human, and subsequently the ratio of unique alleles in the two species was calculated by dividing the n_e of the human by the n_e of the chimpanzee. The median was defined as the 5000th value and the lower and upper confidence limits were defined as the 250th and 9750th values emanating from the bootstrapping. The calculation was performed using the R statistical package (version 2.4.1.) (28). The program MICRO-CHECKER version 2.2.3 was used to check for nonamplified alleles (null alleles) in the microsatellite data set (29) (Table S3 and S4, Supplementary material). The programme MSVAR version 1.3 (www.rubic.rdg.ac.uk/~mab/software) was used to estimate the posterior distribution of different parameters on the chimpanzee MHC and non-MHC microsatellite dataset: N_0 (current population size), N_1 (ancestral population size), t_{fa} (time in years since the population declined or expanded) and μ (mutation rate) with the exponential and linear models. The new version of the MSVAR programme allows checking for locus-specific demographic parameter values, which is interesting when searching for the signature of selection in multilocus data. Markov Chains Monte Carlo (MCMC) of $5*10^9$ updating steps were run with a thinning interval of 50 000 and three runs were performed per demographic model. As a burn-in, the first 10 000 of the 100 000 values of each MCMC sampling were discarded. Starting values for the different demographic parameters as well as their mean and variance were set up as recommended in the MSVAR manual. Convergence and mixing of the chains were evaluated by examining the parameter trends plots, and a generation of 20 years was applied to estimate t_{fa} in years. The programme MODELTEST 3.7 (30) was used to estimate the best substitution model for the mtDNA data, which was (TVM+G). The MODELTEST output was implemented in the PAUP* programme version 4.0b10 for Macintosh (31) to calculate the best tree for the mtDNA data, using maximum likelihood (ML) analysis. Bootstrap values were based on 1000 resamplings.

Results and Discussion

MtDNA analysis: a phylogenetic comparison

The samples were first subjected to mtDNA D-loop analysis. The phylogenetic tree illustrates that the human individuals form, as is known, a highly homogeneous group (Fig. 2). Branch lengths are short, indicating that the observed diversification is of relatively recent origin. The chimpanzees analysed show more diversity, and longer branch lengths are supported by bootstrap values. Indeed, these Great Apes are known to display more mtDNA variation than humans do (32, 33). The different chimpanzee subspecies shared a common ancestor approximately 1.5 Ma (34), whereas the modern human lineage started about 150 000 to 200 000 years ago (35). This indicates that the chimpanzee as a species was subjected to a different time scale to generate variation. This is not only reflected in more mtDNA variation in chimpanzees but also in more variation for some nuclear genes (36, 37). Furthermore, these data show that most *P.t.v.* individuals selected for the present communication are unrelated, display abundant mtDNA variation and as such, represent an outbred population. Although some chimpanzees were found to share identical mtDNAs, such animals possess different MHC haplotypes.

Microsatellite analysis: a comparison between humans and chimpanzees

Markers located at different positions in the MHC region were chosen to define microsatellite profiles. As chimpanzees are an older species than modern humans, one would expect to find in chimpanzees a more diverse, or at least an equivalent complex microsatellite profile. Different randomly selected unlinked autosomal microsatellite markers located outside the MHC region were included to evaluate the genetic variation at the genome scale. For those non-MHC markers both species show about equal amounts of variation (Fig. 3a), and this is also reflected by the expected heterozygosity index (H_E). The average H_E value is 0.63 ± 0.15 for humans and 0.61 ± 0.20 for chimpanzees (Table 1). For most MHC markers, the allelic variation was lower in chimpanzees than in humans (Fig. 3b). The exception is provided by D6S2876 and D6S265, which display equal amounts of allelic variation in both species. However, when the allelic variation is combined with the H_E index, the only outlier seemed to be D6S265. For all other MHC markers tested, humans showed an increased H_E index in comparison to chimpanzees. The average H_E value for the MHC markers is 0.78 ± 0.11 for humans and 0.52 ± 0.27 for chimpanzees (Table 1). The MICRO-CHECKER programme revealed that for the non-MHC markers, both the human and the chimpanzee population analysed are in Hardy–Weinberg equilibrium (HWE), and the loci investigated did not show evidence for null alleles. The same was observed for the MHC markers in the chimpanzee population. This is also applicable to the human population studied, with one exception: marker D6S2876, which showed signs of a null allele.

To verify whether the number of microsatellite alleles detected in humans is significantly higher as compared to that in chimpanzees, the difference in unique number of alleles (Δn_e)

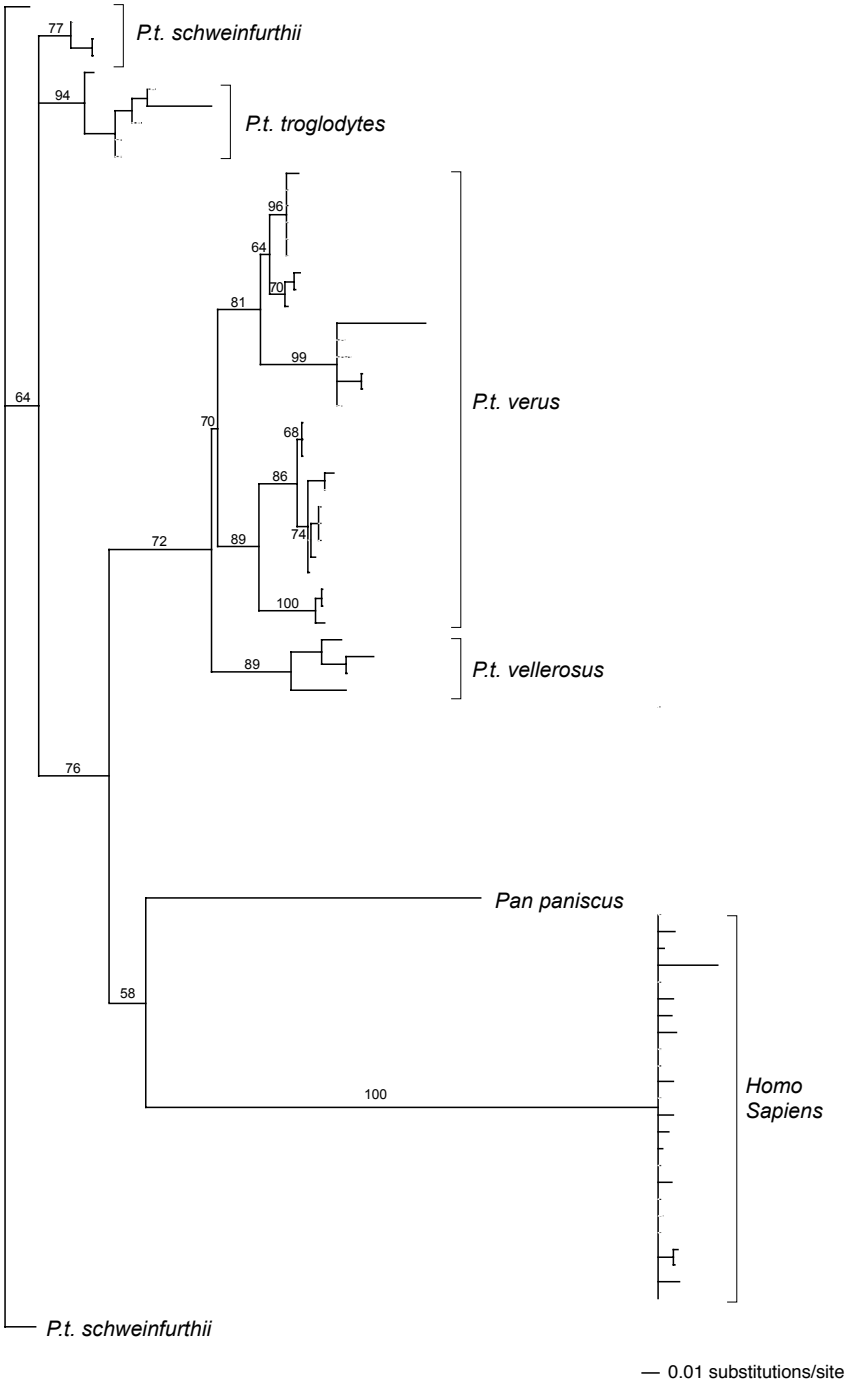


Fig. 2. Unrooted phylogenetic tree of the mtDNA D-loop sequences of human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) samples. The sequence of a *Pan paniscus* (bonobo) is taken as an out-group. Included in the tree are sequences of the other subspecies of chimpanzees (32, 34). Relevant bootstrap values are indicated.

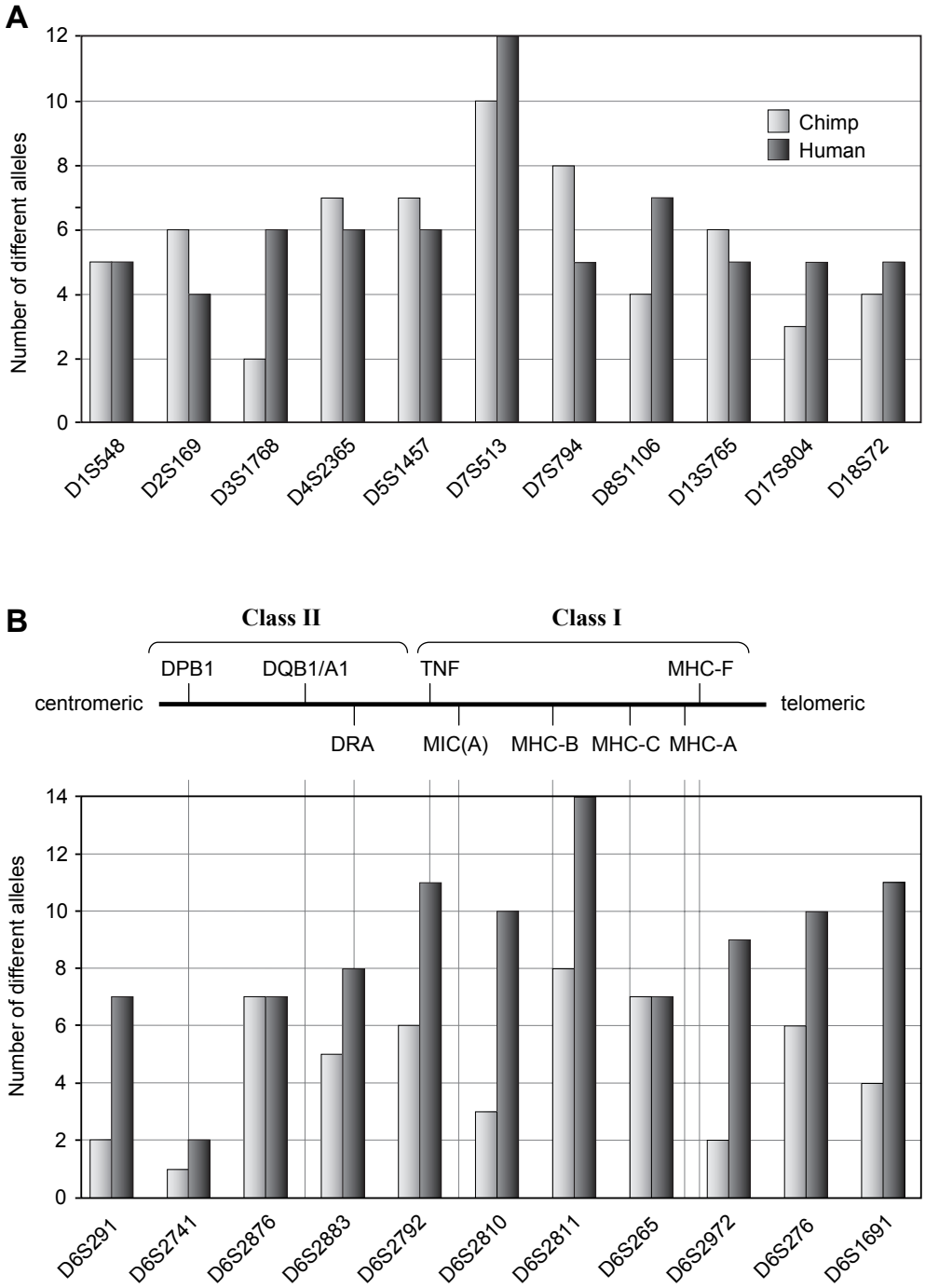


Fig. 3. Bar graphs showing the number of alleles detected for the non-MHC (A) and MHC (B) microsatellite markers analyzed in a panel of 21 chimpanzee individuals compared to 21 human individuals for the non-MHC markers and 25 individuals each for the MHC markers.

was calculated by using bootstrap analyses. Additionally, for each microsatellite, the ratio of the number of unique alleles in humans divided by the number of unique alleles in chimpanzees was calculated. The Δn_e and the ratios of unique alleles were considered statistically significant when their confidence interval (CI) did not enclose 0 or 1, respectively. Nine of the 11 non-MHC markers did not show a statistical significant difference for Δn_e and ratio in humans vs. chimpanzees (Table 1). In contrast, eight of the 11 MHC markers did show a significant different Δn_e value and ratio (Table 1). An analysis of additional human and *P.t.v.* samples did not change the aforementioned results. All extra microsatellite alleles detected in the *P.t.t.* and *P.t.s.* animals analysed are also listed (Table 2).

Although humans and chimpanzees share a high percentage of similarity at the DNA level, it must be taken into account that the presently chosen microsatellites were originally developed for humans, and that length comparison between species can be complicated by ascertainment bias (38-41). The degree of polymorphism seems to be positively correlated with microsatellite length (42). Microsatellite markers are usually longer and more polymorphic in the species from which they were cloned (40). Dinucleotide repeats seem on average to be longer in humans than in chimpanzees, whereas this has not been observed for tetranucleotide repeats (20, 38). To address the issue described above, randomly chosen unlinked autosomal microsatellite markers located outside the MHC region were included in the study. The Δn_e value for the non-MHC markers was equally distributed: sometimes chimpanzees showed more variation ($\Delta n_e \leq -1.00$) and sometimes humans did ($\Delta n_e \geq 1.00$) (Table 1). This was in keeping with the expectations, because at this stage we cannot rule out that other chromosomal segments, apart from Chromosome 6, also experienced selection. On the other hand, for the MHC markers, most of the time more variation was observed for humans than for chimpanzees ($\Delta n_e \geq 1.00$). Moreover, the human microsatellite sequences were compared to their chimpanzee equivalents (Fig. S1, Supplementary material). Slippage of the polymerase increases if the repeat number of the dinucleotide microsatellite is at least eight or higher (43). All dinucleotide microsatellite markers studied had a repeat length of eight or higher in humans as well as in chimpanzees, indicating that in both species there was a greater chance that mutations occur and that microsatellite variation could be generated. In addition, base substitutions within microsatellites are known to reduce the mutation rate (43). Some of the markers studied contain base substitutions. For the markers D6S265, D6S1691 and D6S2883, only the human sequences showed base substitutions (Fig. S1), indicating that for these markers humans probably have a lower mutation rate than chimpanzees. However, in humans these markers show a longer microsatellite allele length than in their chimpanzee equivalents (Table S3, Supplementary material), which may indicate that during chimpanzee evolution variation was lost. There are also dinucleotide markers that showed base substitutions in both species (Fig. S1), indicating that in both species the mutation rate would be reduced. Because chimpanzees are as a species older than humans, we expected to find more variation for the chimpanzee, but this was not observed; hence, the favoured explanation is that this species had lost variation.

In conclusion, the present data show that almost all MHC microsatellite markers studied

Table 1. Total number of alleles detected in the 25 chimpanzees (*Patr*) compared to 25 humans for the different MHC microsatellite markers and for 21 individuals each for the non-MHC markers. The observed heterozygosity (H_o), expected heterozygosity (H_E), the difference in the number of unique alleles [Δn_e (95% CI)] and the ratio of the number of unique alleles in humans divided by the number of unique alleles in chimpanzees (95% CI) were calculated. The microsatellite markers with a significant higher Δn_e and ratio of unique alleles in humans vs. chimpanzees are marked with a grey box.

Non-MHC markers																						
	D1S548		D2S169		**D3S1768**		D4S2365		D5S1457		D7S513		D7S794		**D8S1106**		D13S765		D17S804		D18S72	
	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human
No. of alleles	5	5	6	4	2	6	7	6	7	6	10	12	8	5	4	7	6	5	3	5	4	5
Ho	0.52	0.71	0.67	0.43	0.10	0.76	0.62	0.48	0.81	0.71	0.76	1,00	0.67	0.81	0.57	0.76	0.71	0.52	0.48	0.76	0.48	0.48
He	0.67	0.66	0.71	0.42	0.09	0.78	0.70	0.40	0.77	0.72	0.80	0.84	0.63	0.67	0.66	0.72	0.69	0.69	0.44	0.57	0.59	0.43
Δn_e (95% CI)	0.00 (-2.00 - 1.00)		-2.00 (-3.00 - -1.00)		**4.00 (3.00 - 5.00)**		-1.00 (-3.00 - 1.00)		-1.00 (-3.00 - 1.00)		2.00 (-1.00 - 4.00)		-2.00 (-4.00 - 0.00)		**3.00 (1.00 - 4.00)**		-1.00 (-2.00 - 0.00)		1.00 (0.00 - 2.00)		1.00 (-1.00 - 2.00)	
Ratio (95% CI)	1.00 (0.60 - 1.25)		0.67 (0.50 - 0.80)		**3.00 (2.50 - 6.00)**		0.83 (0.50 - 1.25)		0.83 (0.57 - 1.20)		1.14 (0.88 - 1.57)		0.67 (0.43 - 1.00)		**1.75 (1.25 - 2.33)**		0.83 (0.60 - 1.00)		1.33 (1.00 - 2.00)		1.25 (0.75 - 1.67)	
MHC markers																						
	D6S291		D6S2741		D6S2876		**D6S2883**		**D6S2792**		**D6S2810**		**D6S2811**		D6S265		**D6S2972**		**D6S276**		**D6S1691**	
	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human	Patr	Human
No. of alleles	2	7	1	2	7	7	5	8	6	11	3	10	8	14	7	7	2	9	6	10	4	11
Ho	0.12	0.68	0	0.52	0.64	0.64	0.44	0.76	0.64	0.92	0.52	0.84	0.88	0.92	0.88	0.88	0.56	0.84	0.76	0.68	0.32	0.92
He	0.11	0.69	0	0.49	0.74	0.81	0.40	0.80	0.65	0.87	0.61	0.84	0.82	0.88	0.76	0.74	0.46	0.82	0.70	0.74	0.42	0.88
Δn_e (95% CI)	**4.00 (3.00 - 5.00)**		1.00 (1.00 - 1.00)		0.00 (-2.00 - 1.00)		**3.00 (1.00 - 4.00)**		**5.00 (3.00 - 6.00)**		**6.00 (4.00 - 7.00)**		**5.00 (2.00 - 7.00)**		0.00 (-2.00 - 1.00)		**6.00 (5.00 - 7.00)**		**3.00 (1.00 - 5.00)**		**7.00 (5.00 - 9.00)**	
Ratio (95% CI)	**3.00 (2.50 - 6.00)**		2.00 (2.00 - 2.00)		1.00 (0.71 - 1.17)		**1.60 (1.20 - 2.33)**		**1.83 (1.50 - 2.50)**		**3.00 (2.33 - 3.33)**		**1.63 (1.25 - 2.17)**		1.00 (0.71 - 1.17)		**4.00 (3.50 - 4.50)**		**1.60 (1.17 - 2.33)**		**3.33 (2.25 - 5.50)**	

Table 2. Summary of extra microsatellite alleles detected in the additional analyzed humans and chimpanzees as compared to Table 1. Numbers in brackets indicate the number of individuals analyzed. A dash (–) indicates that no additional microsatellite alleles were detected.

Non MHC markers											
	D1S548	D2S169	D3S1768	D4S2365	D5S1457	D7S513	D7S794	D8S1106	D13S765	D17S804	D18S72
P.t.t. (5)	141	222/224	177/181	280	125/129/137	-	174/182	128/152/156	-	101/103/105	310/314/324
P.t.s. (5)	-	220	-	-	125/129/133	170/192/202	178/182	120	-	101/103/105	314/332
P.t.v. (11)	-	-	-	280/300	-	196	-	-	181/209	99?/103/105	302/324
Human (22)	175	-	-	-	-	188/196/204	160	128	-	112	-
MHC markers											
	D6S291	D6S2741	D6S2876	D6S2883	D6S2792	D6S2810	D6S2811	D6S265	D6S2972	D6S276	D6S1691
P.t.t. (5)	196/198/200	-	218/220/224/246	258/260/266	-	-	94/96/100	109/115/117	-	224	193/195
P.t.s. (5)	198	-	-	256	-	-	94/96	115	126	224	171/179/189/201
P.t.v. (11)	198	247	-	244	-	-	100	117	-	-	195/201
Human (22)	-	-	-	-	-	206/216	107/111	133	-	225	208/226/240/242

display a reduced variation in chimpanzees as compared to humans. This in contrast to the unlinked autosomal markers located outside the MHC region, which showed a more equally distributed genetic variation between both species. The results indicate that, as opposed to humans, the chimpanzee Mhc class I as well as the class II region possesses a reduction in the genetic repertoire.

MHC haplotypes in a P.t.v. population

The pedigreed chimpanzees were also analysed for allelic MHC variation and haplotypes. Nine different *DPB1* alleles were detected, which appear to cluster into two lineages. The *Patr-DPB1*13*, **16* and **17* alleles were frequently observed (Fig. 4a). Variation detected within the *Patr-DPB1* alleles seems to be caused mainly by point mutations (44). At the *HLA-DPB1* locus, variation is caused to some extent by point mutations, but more importantly, a frequent exchange of sequence motifs by recombination-like processes seems to be responsible for generating polymorphism. This may explain, at least in part, the broader range of allelic variation found in humans vs. chimpanzees (45-47). The fact that the chimpanzee alleles are highly related and seem to group only into two lineages minimizes the chance that many novel alleles can be generated by an exchange of polymorphic sequence motifs.

As was shown for humans, particular *Patr-DQA1* and *-DQB* alleles are tightly linked and eight different tandems could be detected (48). The *Patr-DQA1*2002/DQB1*0302* pair was most frequently observed followed by the *Patr-DQA1*0101/DQB1*0602* and *-DQA1*0502/DQB1*0302* tandems (Fig. 4b). However, as compared to humans, chimpanzees show limited allelic variation for the *DQA1* and *DQB1* loci (22, 49), which is also reflected by the presence of a limited combination of *DQA1/DQB1* pairs.

Extensive analysis of the *DRB* region allowed the detection of six different region configurations (Fig. 5). Some previously reported region configurations are modified slightly based on the present results (22). In contrast to earlier reports (50), all six configurations do possess a *DRB1* gene, as is also the case in humans (51). Region configurations I and II were observed most abundantly in the population (Fig. 4c). Although the number of configurations is highly similar in humans and chimpanzees, the *Patr-DRB* loci themselves display limited lineage and allelic variation.

For the *Mhc* class I loci it has been reported that, with regard to numbers, chimpanzees show as much allelic variation as do humans (13). However, chimpanzees lack the equivalents of five out of six *HLA-A* families (13-15). In addition, subsequent intron-2 analysis of the different class I loci has revealed a repertoire reduction that is most prominent for *Patr-B* (18). In humans and chimpanzees, the *Mhc-B* and *-C* loci are closely linked on the chromosome, and the percentage at which particular *Patr-B/C* pairs were observed in the panel is shown (Fig. 4d). The *Patr-B*0101/C*0401* pair was present in most abundance, whereas three other pairs, *-B*0501/C*0601*, *-B*1401/C*0203* and *-B*2402/C*0901*, were moderately present. The different *Patr-A* alleles detected in the population studied seemed to be more equally distributed among the different haplotypes, although the presence of *-A*0301*, *-A*0401* and *-A*0901* seemed to be slightly increased (Fig. 4e).

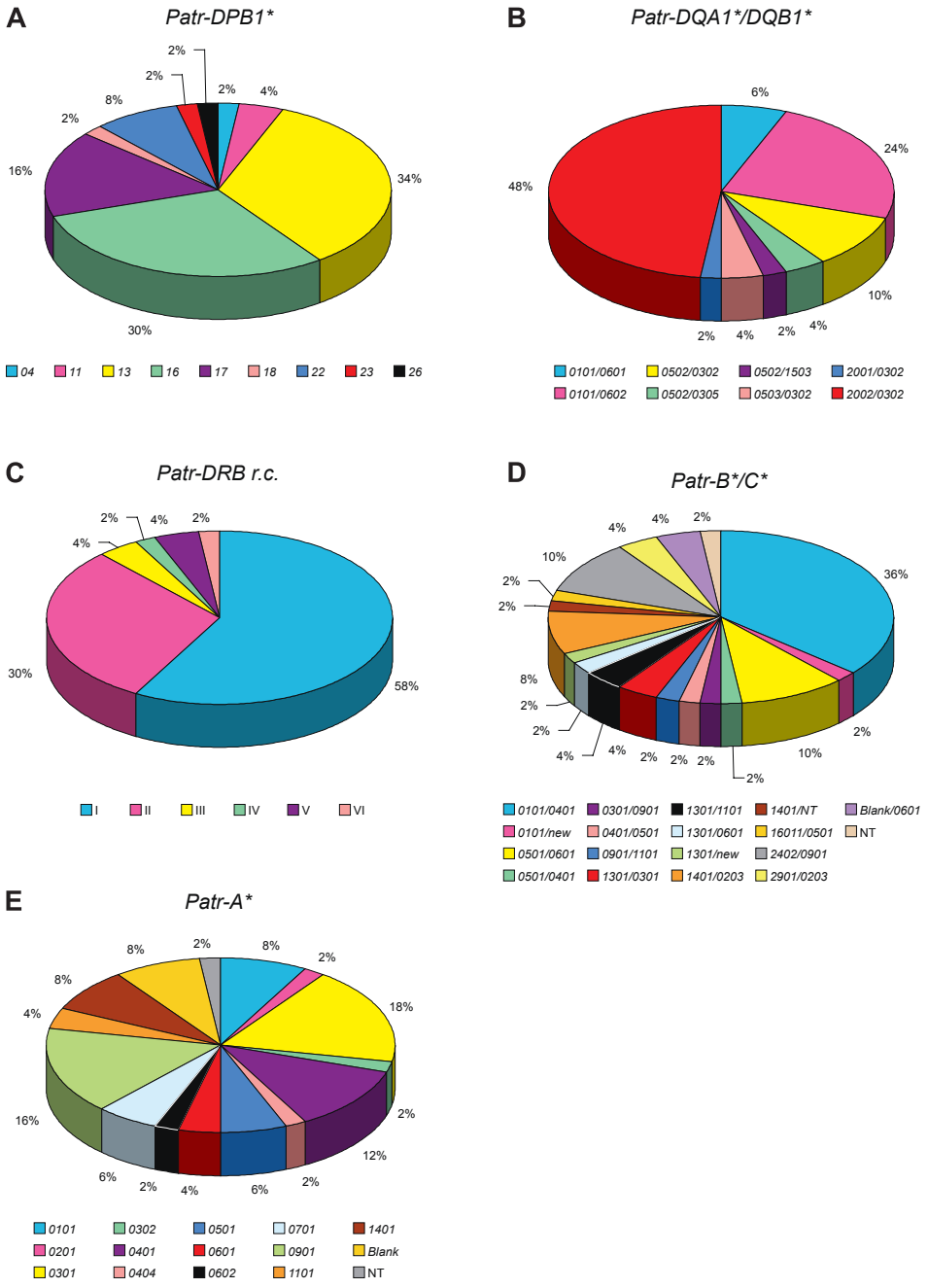


Fig. 4. Pie charts showing the percentage at which a *Mhc* allele, allele combination or region configuration was present in the chimpanzee population studied. The figure shows the frequencies of (A) *Patr-DPB1** alleles, (B) *Patr-DQA1*/DQB1** pairs, (C) *Patr-DRB* region configurations (r.c.), (D) *Patr-B*/C** tandems and (E) *Patr-A** alleles. New, indicates a newly detected allele that has not yet received an official designation. Blank indicates that the MHC molecule is absent based on serology and reference strand conformation analyses (RSCA) (14). NT indicates that the particular allele was not typed.

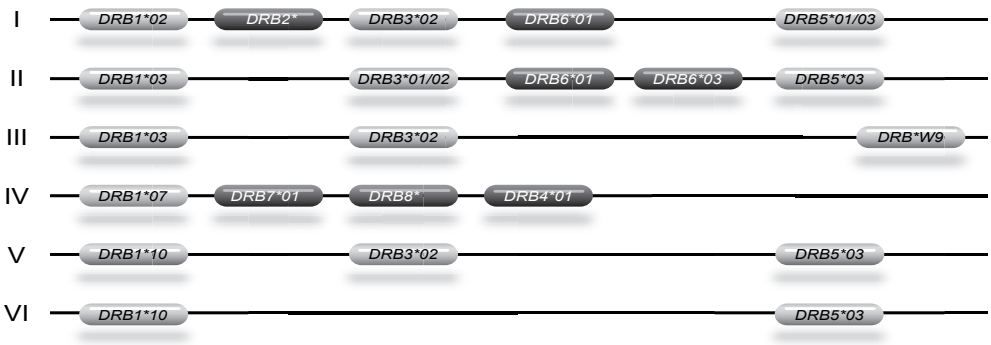


Fig. 5. *Patr-DRB* region configurations. For instance, in configuration I, *DRB5*01/03* means that lineage polymorphism is detected for this locus. Configuration III possesses a locus that received a Workshop number (W). Configuration IV is shared between humans and chimpanzees. Black boxes represent pseudogenes.

Based on the abundant presence of the *Patr-B*0101/C*0401* tandem (36%), it is possible that this allelic combination may have had an important immunological function in the evolutionary past of chimpanzees. MHC class I molecules play an important role in binding peptides, and different allotypes have distinct peptide binding pockets (52-54). For instance, the anchoring binding pockets of the Patr-B*0101 and HLA-B*5701 molecules are highly similar, indicating that both molecules are able to present identical or highly related peptides to CTLs. The HLA-B*5701 molecule is associated with protection against disease progression in HIV-1 exposed or infected human individuals (55). HIV-1/SIV_{cpz} infected chimpanzees are known to be relatively resistant to develop AIDS. Positive selection for the *Patr-B*0101/C*0401* tandem in the past, based on its ability to protect against disease progression, could be an explanation for the abundant presence of this tandem in the contemporary population. Alternatively, the appearance of a particular allele/haplotype at high frequency could also be explained by a founder effect or genetic drift. However, the observation that the small population studied possesses considerable mtDNA variation argues against a founder effect. Furthermore, whereas the studied population contained only a limited amount of *Mhc* lineages, some of these lineages appeared to show allelic variation. This observation argues against the possibility that a particular haplotype is favoured due to genetic drift, and therefore the most likely origin of the abundant presence of the *Patr-B*0101/C*0401* tandem would be a selective sweep. In addition, we examined whether particular *Patr-B/C* pairs were preferentially linked to certain *-A* alleles or vice versa. The *Patr-A* locus is located some 1300 kb telomeric of *-B/C* tandem on the chromosome. Unique combinations were rare, indicating that recombination occurs between these chromosomal segments. There are particular *Patr-B/C* tandems found in combination with only one *-A* allele (Fig. S2A, Supplementary material),

but these *-A* alleles were detected in combination with other *-B/C* tandems. Furthermore, no strict pairing was observed between the *Patr-B/C* pairs and the different class II loci (Fig. S2B-D), which are separated by the *Mhc* class III region, spanning about 700 kb. Moreover, the abundantly present *Patr-B*0101/C*0401* pair was detected in combination with many different *-A* and *-DPB1* alleles, as well with different, mainly highly present, *Patr-DQA1/DQB1* pairs or *-DRB* region configurations. Finally, the data shows that if, for instance, selection took place today on the *Patr-B*0101/C*0401* pair, considerable variation would be selected with this tandem. On the other hand, if selection were to take place on the most abundantly present *Patr-DQA1/DQB1* pair or *-DRB* region configuration, a great deal of *-B/C* variation would be selected. The examples show that in spite of selection for a particular molecule or tandem, it is possible that sufficient variation will remain in a population.

The haplotype analysis shows that the chimpanzee *Mhc* class I and class II loci possess a reduction in the repertoire as compared to humans. The observation corroborates with the microsatellite data, where almost all markers located in the MHC region showed a reduced variation in chimpanzees in relation to the human data. Furthermore, the reduced genetic variation in the chimpanzee MHC is best ascribed to a selective sweep, and a founder effect or genetic drift could be excluded due to the presence of considerable mtDNA variation and allelic variation, respectively. Finally, most of the 25 animals studied possess unique haplotypes. This was quite unexpected, since the animals were sampled from one population originating from Sierra Leone. It is most likely that some of the haplotypes were selected positively, and after the selective sweep naturally occurring recombination-like processes contributed in recovering *Mhc* haplotype diversity.

Defining the MHC region segment most prominently targeted by the selective sweep

Mhc class I intron analyses indicated that mainly the *Patr-B* locus repertoire was targeted by the selective sweep (18). The microsatellite data illustrate that marker D6S2810 [Δn_e = 6.00 (4.00 – 7.00), Ratio = 3.00 (2.33 – 3.33)], located between *MIC* and the *Mhc-B* loci, is significantly reduced in chimpanzees. In humans, this marker shows linkage disequilibrium with *HLA-B* (56). In the chimpanzee population, three D6S2810 alleles were detected (Table 1) from which alleles 190 and 192 were often found in combination with the *Patr-B*0101/C*0401* haplotype, though allele 190 was also found in combination with the *-B*2402/C*0901* haplotype. The 188 allele segregated with all other *Patr-B/C* haplotypes detected in the population. The microsatellite marker D6S2811 [Δn_e = 5.00 (2.00 – 7.00), Ratio = 1.63 (1.25 – 2.17)], located between the *Mhc-B* and *-C* loci, showed evidence for a repertoire reduction in chimpanzees as well. These two markers, located near the *Mhc-B* locus, and the observation that six of the seven analyzed markers located in or close to the *Mhc* class I region were showing a reduced variation in chimpanzees vs. humans (Table 1), support the results published for the *Mhc* class I and major histocompatibility complex class I chain-related gene (*MIC*) analyses (18, 23).

The ancient selective sweep in the chimpanzee *Mhc* class I region was dated to have

happened approximately 2-3 Ma (18). However, it is difficult to date an ancient selective sweep with microsatellite data due to their high mutation rate. The MSVAR multilocus analysis of the MHC and non-MHC microsatellite data showed that, whatever the demographic model assumed: linear or exponential, the chimpanzee population had decreased from at least 10 000 to around 100 effective numbers of individuals (Fig. 6). This decrease seems to be recent: for the exponential model >3000 and <10 000 years, and for the linear model between 10 000 and 100 000 years. Figure 6 further shows that, assuming a linear model of population decrease, the MHC data suggest an older decrease of variability than do the non-MHC data, which may reflect a signature of the ancient selective sweep in the *Mhc* class I gene repertoire of chimpanzees. Taking into consideration the MSVAR single locus estimation (Fig. 7), the variance for the MHC and non-MHC markers was very similar but not evenly distributed. The non-MHC markers seemed to have a variance that reflects the genealogical stochasticity, whereas the MHC markers could be divided into two groups. Five markers were located in the *Mhc* class I region with a log (N_1) of approximately five and four markers in the *Mhc* class II region with a log (N_1) of approximately four. Since the current population size (N_0) is the same for all MHC markers, the data suggest that it is more likely that genetic variation in the *Mhc* class I region has been reduced mostly by selection.

Furthermore, the MHC haplotype data illustrate that other genes, apart from the class I *B* genes, show a reduced repertoire as well. The *HLA-A* locus alleles can be divided into an *HLA-A2* and *-A3* lineage, based on the substitution of 33 diagnostic nucleotide positions (57, 58). Chimpanzees possess only orthologues that belong to the *HLA-A3* lineage (15). However, orthologues of the *HLA-A2* lineage are described in gorillas (59), indicating that this lineage predates the speciation of humans and gorillas. Moreover, chimpanzees lost the equivalents of the *HLA-DRB1*04* lineage, whereas exon-2 sequences grouping in this lineage are present in Old World monkeys (60). These examples demonstrate that chimpanzees indeed lost particular *Mhc* lineages during evolution.

In addition, other loci within the chimpanzee MHC region seem to be affected as well and show signs of repertoire reduction, as in the case of the *MIC* gene, located in the class I region centromeric of the *Mhc-B* locus. Humans have two copies: namely, a *MICA* and a *MICB* gene, whereas the chimpanzee *MIC* locus appears to be a fusion product of the two loci (61), comprising one lineage showing moderate allelic variation (23). In the chimpanzee panel, only three alleles could be detected. Orthologues of the *MICA* and *MICB* genes have been found in rhesus macaques, orang-utans and humans (62-64). This indicates that haplotypes with both genes were present in a common ancestor. A gene fusion took place in progenitors of contemporary chimpanzees, but haplotypes carrying the separated *MICA* and *MICB* genes were lost (23). Another example is provided by the *C4* gene, a component of the complement system and mapping in the central MHC region. Most humans possess two copies, designated *C4A* and *C4B*, which are known to differ in size. The long version of the *C4A* gene, caused by an ancient retroviral insert, is present in humans and orang-utans. Chimpanzees, however, possess only the short version of the *C4A* and *C4B* genes (65). As the presence of the *C4A* long version predates the speciation of humans and orang-utans,

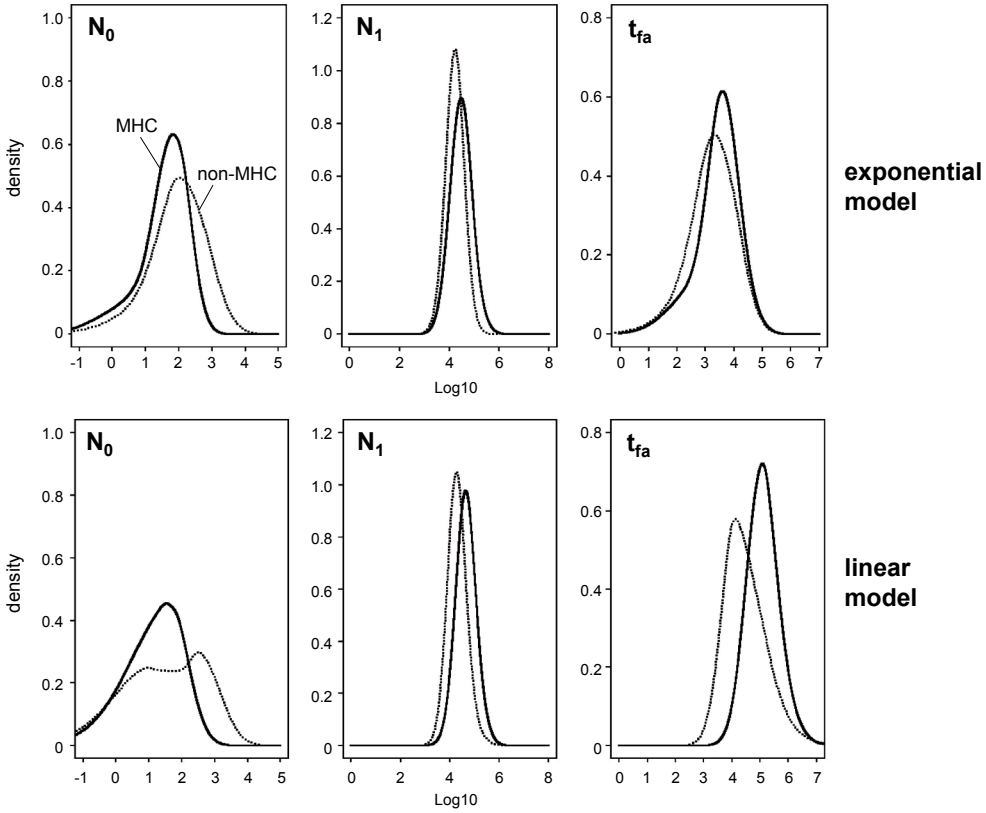


Fig. 6. Multilocus demographic parameters estimation for the chimpanzee MHC (solid line) and non-MHC (dotted line) microsatellite data using the exponential and linear model. N_0 stands for current population size, N_1 for ancestral population size and t_{fa} for time in years since the population declined or expanded.

it is most likely that this version was lost during chimpanzee evolution.

The examples described above illustrate that the entire chimpanzee MHC region is showing various signs of an allelic repertoire reduction, and that particular lineages or gene moieties have been lost. Nevertheless, combining the results of the microsatellite analysis and the MHC haplotypes indicates that the selective sweep acted most prominently on the *Mhc* class I region, and that particular *Mhc* class II alleles and other genes were selected based on the fact that they were genetically coupled to particular class I region alleles. As can been seen, a strong selection on the highly frequent *Patr-B*0101/C*0401* tandem would, even currently, result in many *Patr-A* alleles/lineages and *Mhc* class II loci/lineages being rescued (Fig. S2, Supplementary material). New haplotypes were generated by recombination, and this phenomenon may erase the signal of the selective sweep over generations at the haplotype level. However, the multilocus demographic analyses illustrate that five MHC class I loci show

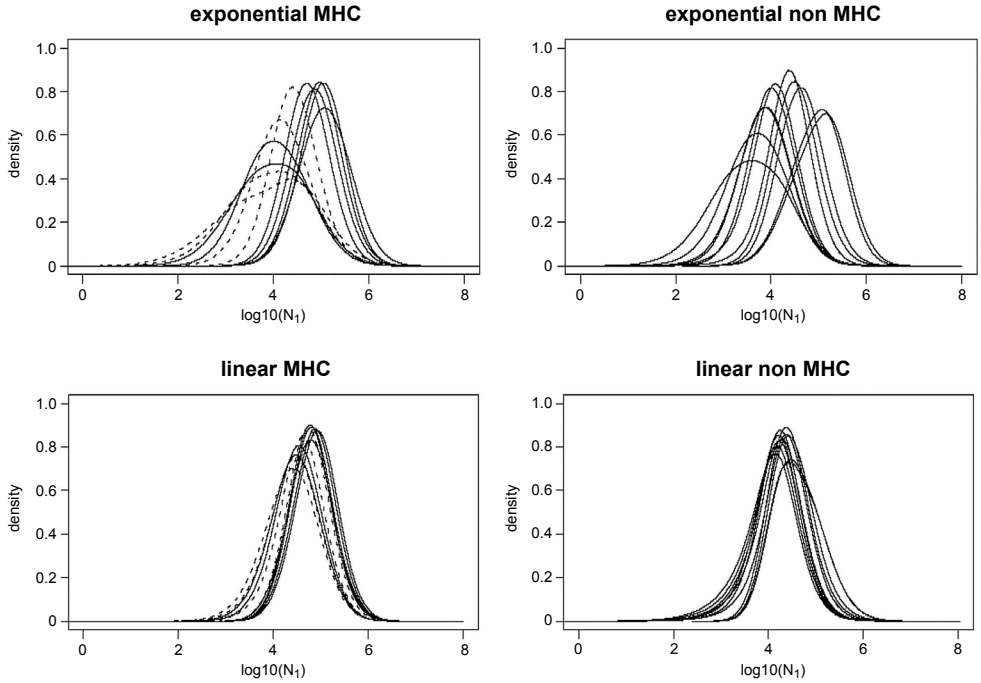


Fig. 7. Single-locus estimations for the chimpanzee MHC and non-MHC microsatellite data using the exponential and linear model. N_1 stands for the ancestral population size. For the MHC graphs the solid lines represent the MHC class I microsatellites, whereas the dotted lines represent the MHC class II microsatellite markers.

the same pattern of strongly reduced genetic variation. Genetic data from these loci on the same chromosomal segment thus indicate that the selective sweep has been strong and that recombination did not erase the signal.

Acknowledgements

The authors wish to thank Dr M. Carrington for critically reading the manuscript, D. Devine for editing the manuscript, H. van Westbroek for preparing the figures, Drs F. Claas and I. Doxiadis for providing the human DNA samples and G. Adam for technical assistance. This study was supported in part by NIH/NIAID contract number HHSN266200400088C and 5R24RR016038-05

References

1. Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67-74.
2. Brown JH, Jardetzky T, Saper MA, et al. (1988) A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature 332:845-850.
3. Bjorkman PJ, Saper MA, Samraoui B, et al. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512-518.
4. Fujiyama A, Watanabe H, Toyoda A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
5. Bergstrom TF, Engkvist H, Erlandsson R, et al. (1999) Tracing the origin of *HLA-DRB1* alleles by microsatellite polymorphism. Am J Hum Genet 64:1709-1718.
6. Bontrop RE (2006) Comparative genetics of MHC polymorphisms in different primate species: duplications and deletions. Hum Immunol 67:388-397.
7. von Salome J, Gyllensten U, Bergstrom TF (2007) Full-length sequence analysis of the *HLA-DRB1* locus suggests a recent origin of alleles. Immunogenetics 59:261-271.
8. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287:607-614.
9. Santiago ML, Rodenburg CM, Kamenya S, et al. (2002) SIVcpz in wild chimpanzees. Science 295:465.
10. Heeney JL, Rutjens E, Verschoor EJ, et al. (2006) Transmission of simian immunodeficiency virus SIVcpz and the evolution of infection in the presence and absence of concurrent human immunodeficiency virus type 1 infection in chimpanzees. J Virol 80:7208-7218.
11. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
12. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, et al. (1999) Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol 162: 2308-2314.
13. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. Immunogenetics 51:410-424.
14. de Groot NG, Otting N, Arguello R, et al. (2000) Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. Immunogenetics 51:398-409.
15. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
16. Bodmer WF (1972) Evolutionary significance of the HL-A system. Nature 237:139-145 passim.
17. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.
18. de Groot NG, Otting N, Doxiadis GG, et al. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci U S A 99: 11748-11753.
19. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
20. Webster MT, Smith NG, Ellegren H (2002) Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. Proc Natl Acad Sci U S A 99:8748-8753.

21. Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160:753-763.
22. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
23. de Groot NG, Garcia CA, Verschoor EJ, et al. (2005) Reduced *MIC* gene repertoire variation in West African chimpanzees as compared to humans. Mol Biol Evol 22:1375-1385.
24. Gourraud PA, Mano S, Barnetche T, et al. (2004) Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. Tissue Antigens 64:543-555.
25. Penedo MC, Bontrop RE, Heijmans CM, et al. (2005) Microsatellite typing of the rhesus macaque MHC region. Immunogenetics 57:198-209.
26. Khazand M, Peiberg C, Nagy M, Sauermann U (1999) Mhc-DQ-DRB haplotype analysis in the rhesus macaque: evidence for a number of different haplotypes displaying a low allelic polymorphism. Tissue Antigens 54:615-624.
27. Doxiadis GG, Otting N, de Groot NG, et al. (2000) Unprecedented polymorphism of *Mhc-DRB* region configurations in rhesus macaques. J Immunol 164:3193-3199.
28. R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical computing, Vienna, Austria. http://www.r-project.org/ (last checked 08-Feb-2008)
29. van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.
30. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.
31. Swofford DL (1992) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods).
32. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
33. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713.
34. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
35. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc Natl Acad Sci U S A 103:9381-9386.
36. Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.
37. Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A 97:11354-11358.
38. Cooper G, Rubinsztein DC, Amos W (1998) Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. Hum Mol Genet 7:1425-1429.
39. Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. Hum Mol Genet 5:1131-1137.
40. Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nat Genet 11:360-362.
41. Rubinsztein DC, Leggo J, Amos W (1995) Microsatellites evolve more rapidly in humans than in chimpanzees. Genomics 30:610-612.

42. Kayser M, Vowles EJ, Kappei D, Amos W (2006) Microsatellite length differences between humans and chimpanzees at autosomal Loci are not found at equivalent haploid y chromosomal Loci. Genetics 173:2179-2186.
43. Vowles EJ, Amos W (2006) Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences. Mol Biol Evol 23:598-607.
44. Otting N, Doxiadis GG, Versluis L, et al. (1998) Characterization and distribution of *Mhc-DPB1* alleles in chimpanzee and rhesus macaque populations. Hum Immunol 59:656-664.
45. Gyllensten U, Bergstrom T, Josefsson A, et al. (1996) Rapid allelic diversification and intensified selection at antigen recognition sites of the Mhc class II DPB1 locus during hominoid evolution. Tissue Antigens 47:212-221.
46. Reinders J, Rozemuller EH, van Gent R, et al. (2005) Extended HLA-DPB1 polymorphism: an RNA approach for HLA-DPB1 typing. Immunogenetics 57:790-794.
47. Slierendregt BL, Otting N, Kenter M, Bontrop RE (1995) Allelic diversity at the *Mhc-DP* locus in rhesus macaques (Macaca mulatta). Immunogenetics 41:29-37.
48. Begovich A, Klitz W, Steiner L, et al. (2000) in Major Histocompatibility complex, ed. Kasahara, M. (Springer Verlag, Tokyo), pp. 412-426.
49. Otting N, de Groot NG, Doxiadis GG, Bontrop RE (2002) Extensive *Mhc-DQB* variation in humans and non-human primate species. Immunogenetics 54:230-239.
50. Slierendregt BL, Kenter M, Otting N, et al. (1993) Major histocompatibility complex class II haplotypes in a breeding colony of chimpanzees (*Pan troglodytes*). Tissue Antigens 42:55-61.
51. Robinson J, Waller MJ, Parham P, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31:311-314.
52. McKinney DM, Erickson AL, Walker CM, et al. (2000) Identification of five different Patr class I molecules that bind HLA supertype peptides and definition of their peptide binding motifs. J Immunol 165:4414-4422.
53. Sidney J, Asabe S, Peters B, et al. (2006) Detailed characterization of the peptide binding specificity of five common Patr class I MHC molecules. Immunogenetics 58:559-570.
54. Sidney J, Peters B, Moore C, et al. (2007) Characterization of the peptide-binding specificity of the chimpanzee class I alleles A*0301 and A*0401 using a combinatorial peptide library. Immunogenetics 59:745-751.
55. Kaslow RA, Carrington M, Apple R, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2:405-411.
56. Grimaldi MC, Clayton J, Pontarotti P, et al. (1996) New highly polymorphic microsatellite marker in linkage disequilibrium with HLA-B. Hum Immunol 51:89-94.
57. Domena JD, Hildebrand WH, Bias WB, Parham P (1993) A sixth family of *HLA-A* alleles defined by *HLA-A*8001*. Tissue Antigens 42:156-159.
58. Kato K, Trapani JA, Allopenna J, et al. (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 143:3371-3378
59. Lawlor DA, Warren E, Taylor P, Parham P (1991) Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. J Exp Med 174:1491-1509.
60. Bontrop RE, Otting N, Slierendregt BL, Lanchbury JS (1995) Evolution of major histocompatibility complex polymorphisms and T-cell receptor diversity in primates. Immunol Rev 143:33-62.

61. Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122.
62. Bahram S, Bresnahan M, Geraghty DE, Spies T (1994) A second lineage of mammalian major histocompatibility complex class I genes. Proc Natl Acad Sci U S A 91:6259-6263.
63. Daza-Vamenta R, Glusman G, Rowen L, et al. (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Research 14:1501-1515.
64. Pellet P, Vaneensberghe C, Debre P, et al. (1999) MIC genes in non-human primates. European Journal of Immunogenetics 26:239-241.
65. Bontrop RE, Broos LA, Otting N, Jonker MJ (1991) Polymorphism of C4 and CYP21 genes in various primate species. Tissue Antigens 37:145-151.
66. Gonder MK, Oates JF, Disotell TR, et al. (1997) A new west African chimpanzee subspecies? Nature 388:337.

Supplementary material

The following supplementary material is available for this article:

Fig. S1. Comparison of the microsatellite sequence (including 20 nucleotides in front and 20 nucleotides after the microsatellite repeat) between humans and chimpanzees for the different MHC and non-MHC markers. A dash indicates identity to the human sequence. A dot indicates a deletion/insertion. Marker D6S2792 comprises two distinct repeats, designated A and B.

Fig. S2. Bar diagrams showing the percentage to what extent a specific A) *Patr-A** allele, B) *Patr-DPB1** allele, C) *Patr-DQA1*/DQB1** pair or D) *Patr-DRB* region configuration (r.c.) is detected in combination with a particular *Patr-B*/C** tandem. For a clear figure only the *Patr-B*/C** tandems with a presence of 4% or higher (Fig. 4) are incorporated. *Blank* indicates that the MHC molecule is absent based on serology and RSCA (de Groot et al. 2000).

Table S1. Information on primers for the MHC and non-MHC microsatellite markers used, the chromosome where the marker is located (Chr. nr.), the fluorescence label, the primer sequence and concentration (Primer conc.) and the source reference. The non-MHC markers are assembled in five different multiplex primer mixtures designated MuA, MuBC, MuD, MuE and MuF. The markers indicated with # were not analyzed in the present study. [References, 1, (Worwood et al. 1994), 2, (Penedo et al. 2005), 3, (Jongeneel et al. 1991), 4, (Grimaldi et al. 1996), 5, (Tamiya et al. 1998), 6, (Andrade et al. 2004), 7: (Penedo, personal communication).]

Table S2. Information on the primers, the annealing temperature (ann. temp) and the $MgCl_2$ concentration or the *DPB1*, *DQA1* and *DQB1* loci.

Table S3. Input file for the MICROSATELLITE ANALYSER (MSA) programme (Dieringer & Schlotterer 2003) of the MHC microsatellite markers analysed in the chimpanzee and human samples. The input file is in the format of a Word text-only document. The marker D6S265 possess an allele that is not in conformation with the normal repeat length, and probably possess a mutation. However, the allele is observed in different chimpanzees, which supports the existence of the allele.

Table S4. Input file for the MSA programme (Dieringer & Schlotterer 2003) of the unlinked autosomal microsatellite markers located outside the MHC region analysed in the chimpanzee and human samples. The input file is in the format of a Word text-only document. Two markers, D1S548 and D7S794, possess an allele that is not in conformation with the normal repeat length. Probably these alleles possess a mutation. For D7S794, segregation within a chimpanzee family is observed, which supports the existence of the particular allele, whereas for D1S548 the allele was observed in different human individuals.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03716.x
(This link will take you to the article abstract).

Chapter 6

The chimpanzee *Mhc-DRB* region revisited: Gene content, polymorphism, pseudogenes, and transcripts

Natasja G. de Groot, Corrine M.C. Heijmans[#], Nanine de Groot[#], Gaby G.M. Doxiadis, Nel Otting, Ronald E. Bontrop
[#]Both authors contributed equally to this work.

Molecular Immunology 47 (2009), 381-389

Department of Comparative Genetics and Refinement, Biomedical Primate Research Centre, Rijswijk, The Netherlands

Abstract

In humans, great apes, and different monkey species, the major histocompatibility complex (MHC) class II *DRB* region is known to display considerable copy number variation. The microsatellite D6S2878 has been shown to be a valuable marker for haplotyping the *DR* region in humans and macaque species. The present report illustrates that chimpanzee haplotypes also can be discriminated with this marker. The analyses resulted in the description of nine different region configurations, of which seven are present within the West African chimpanzee population studied. The region configurations vary in gene content from two up to five *DRB* genes. Subsequent cDNA sequencing increased the number of known full-length *Patr-DRB* sequences from 3 to 32, and shows that one to three *Patr-DRB* genes per haplotype apparently produce functional transcripts. This is more or less comparable to humans and rhesus macaques. Moreover, microsatellite analysis in concert with full-length *DRB* gene sequencing showed that the *Patr-DRB*W9* and *-DRB3*01/02* lineages most likely arose from a common ancestral lineage: hence, the *Patr-DRB*W9* lineage was renamed to *Patr-DRB3*07*. Overall, the data demonstrate that the D6S2878 microsatellite marker allows fast and accurate haplotyping of the *Patr-DRB* region. In addition, the limited amount of allelic variation observed at the various *Patr-DRB* genes is in agreement with the fact that chimpanzees experienced a selective sweep that may have been caused by an ancient retroviral infection.

Introduction

Chimpanzees (*Pan troglodytes*) are the closest living relative of humans, and both species share approximately 98.7% similarity at the nonrepetitive DNA level (1). The split between the ancestors of humans and chimpanzees happened about 5-6 million years ago (1). Chimpanzees can be divided into four different subspecies, designated as *Pan troglodytes troglodytes* (*P.t.t.*), *P.t. schweinfurthii* (*P.t.s.*), *P.t. verus* (*P.t.v.*), and *P. t. vellerosus* (2, 3).

The major histocompatibility complex (MHC) is a multigene family, which is present in most vertebrate species, and two clusters of cell-surface proteins are recognised. MHC class I gene products are expressed on nucleated cells, whereas the MHC class II proteins are expressed on B cells, macrophages, and other antigen-presenting cells. The MHC plays a key role in generating adaptive immunological responses, and the main function is the binding of pathogen-derived peptides and presenting them on the cell surface for T-cell recognition. The MHC class II gene products in humans and chimpanzees are designated *HLA-DP*, *-DQ*, and *-DR*, and *Patr-DP*, *-DQ*, and *-DR*, respectively. All three loci possess at least one *A* gene and one *B* gene. The *DR* region is the most complex, and several *B* genes are characterised in humans: they have been named *HLA-DRB1* to *-DRB9* (4). In chimpanzees, the orthologues of these genes are described, and as in humans, the *-DRB2* and *-DRB6* to *-DRB9* are considered to be pseudogenes (5). Chimpanzees, however, seem to have lost during their evolution

the equivalent of the *HLA-DRB1*04* lineage. In both species, the various types of *DRB* loci display different levels of polymorphism, and most of this variation is confined to exon 2. Additionally, the unique number of *DRB* genes present per region configuration differs and ranges from one to four in humans and from two to five in chimpanzees (6, 7). Allelic variation is observed within region configurations, and as a consequence different haplotypes can be discriminated per region configuration.

A complex dinucleotide repeat, D6S2878, located in close proximity of exon 2 of most *HLA-DRB* genes (8-10), was found to represent an accurate and valuable marker for haplotyping humans and two macaque species (7, 11). The genetic profile of this microsatellite showed its usefulness in the description of *Patr-DRB1* alleles (12). In humans, the *HLA-DRB1*, *-DRB3*, *-DRB4*, and *-DRB5* genes produce functional transcripts that are translated into products. To date there are limited data available on the actual transcription of *Patr-DRB* alleles (13), and even a thorough chimpanzee population study is lacking.

Chimpanzees are known to have very effective cytotoxic T-cell (CTL) responses against human immunodeficiency virus type 1 (HIV-1) infection (14). Moreover, like particular human individuals, chimpanzees in general are relative resistant to developing acquired immunodeficiency syndrome (AIDS) after natural or experimental infection with HIV-1 or chimpanzee-derived simian immunodeficiency virus (SIV_{cpz}, the closest relative of HIV-1) (15-17). In humans, resistance to AIDS is strongly associated with the presence of the HLA-B*2705 or -B*5701 molecules (18); also in chimpanzees CTL responses are found that are directed against conserved HIV-1 epitopes in the context of particular MHC class I molecules (19). Based on these observations, and on the fact that chimpanzees possess a reduced *Mhc* class I repertoire, we hypothesised that the *Mhc* class I gene repertoire reduction may have been caused by SIV_{cpz} or a related ancestral retrovirus. As such, the contemporary chimpanzee populations have modified MHC class I repertoires but are able to cope with their natural environment and with (retro)viral infections such as HIV-1/SIV_{cpz} (20). Evidence for such a selective sweep was further substantiated by the analyses of the *MIC* locus, which maps near *Patr-B* (21), and by comparative genomics using different microsatellite markers mapping inside and outside the MHC region (6).

MHC class II molecules play a pivotal role in providing $CD4^+$ T-cell-mediated help and antibody production. This effector function is also important for the development and maintenance of $CD8^+$ memory T cells (22). Although the MHC class I repertoire of chimpanzees is reduced, these animals do mount effective CTL responses against various pathogens such as HCV, HIV-1, and malaria (14, 23, 24). We aimed to characterise the *Mhc-DRB* region of a *P.t.verus* colony thoroughly and to investigate which different *Patr-DRB* alleles may play a role in providing immunological help to $CD8^+$ memory T cells.

Material and methods

Source of RNA and DNA

Thirty-five chimpanzees originating from Sierra Leone represent the founder population that is the basis of the West African chimpanzee colony studied. The colony increased to more than 200 animals, covering three generations. The offspring was pedigreed based on serological specificities (Patr-A and -B) and molecular-defined class I and II polymorphisms (25, 26). Epstein-Barr virus- transformed B-cell lines were used to obtain RNA and genomic DNA (gDNA). Human DNA samples, originating from unrelated Caucasoid individuals, were provided by the department of Immunohaematology and Blood Bank of the Leiden University Medical Centre (7) and were included for comparison. In addition, samples of four *P.t.t.* and two *P.t.s.* animals were included in the panel.

Patr-DRB short tandem repeat (STR) genotyping and sequencing

The microsatellite marker D6S2878 was used for *DRB* genotyping. The genotyped cohort comprised of 114 different chimpanzee DNA samples of founder animals and offspring. The relevant DNA segment was amplified with the primers (0.2 µM) described for the human samples and the genotyping techniques were performed as published (7, 11).

Forty-three (31 *P.t.v.* and 12 *P.t.t./P.t.s.*) different *Patr-DRB* alleles were sequenced from exon 2 to intron 2, and this area included the microsatellite. The same primers are used as described for humans (7). The PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen) and were subsequently blunt-end ligated into the vector pJET1.2/blunt vector (Fermentas, St. Leon-Rot, Germany). Transformation of the XL1Blue *Escherichia coli* cells was performed with the TransformAid Bacterial Transformation Kit (Fermentas). A minimum of 32 clones were examined per PCR reaction. The sequence samples were prepared using 1 µl of ABI Prism BigDye Terminator v3.1 Cycle Sequencing mixture (Applied Biosystems, Foster City, CA), 0.2 µM of pJET1.2 forward or reverse primer (Fermentas), and 2 µl of 5x sequencing dilution buffer (400mM Tris-HCL, 10mM $MgCl_2$) in a 10 µl reaction, and sequenced on an ABI 3130xl genetic analyser (Applied Biosystems). The data were analysed using the programmes Lasergene SeqMan Pro version 7.2.1 (Dnastar, Inc Madison, USA) and MacVector version 10.0.2. (MacVector, Inc Cambridge, UK). At least two independent PCR reactions were performed and/or the alleles were confirmed by their presence in different animals.

Patr-DRB cDNA sequencing

For 21 different animals (17 *P.t.v.* and 4 *P.t.t.*), covering most of the genotyped *Mhc-DRB* haplotypes, RNA was extracted using the RNeasy mini kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. An RT-PCR reaction (Promega, Madison, WI, USA) in accordance with the manufacturer's recommendation was performed using *Mhc-DRB* specific primers described previously (27). The purification of the PCR products, the ligation, transformation, and sequencing was performed as described above.

Phylogenetic analysis and detection of transposable elements

Neighbor-Joining (NJ) trees were constructed with the PAUP* programme version 4.0b10 for Macintosh (28). Pairwise distances were calculated using the Kimura-2 parameter method. Bootstrap values were calculated based on 1000 replicates. Alleles of the *Patr-DRB3*01/02* and *-DRB*W9* lineages were analysed for their repetitive elements using the CENSOR software tool from the Genetic Information Research Institute website (http://www.girinst.org/censor) (29).

Nomenclature and re-designation of the Patr-DRB*W9 lineage

Allele names were assigned according to a standardized protocol (30-32). The full-length *DRB* sequences are deposited at the EMBL database, and received accession numbers FN424191–FN424222. The newly described exon 2 sequence *Patr-DRB1*0215* received accession number FN424223. All sequences are documented at the IPD-MHC-NHP-database (www.ebi.ac.uk/ipd/mhc/nhp).

Based on gDNA and cDNA sequences, the *Patr-DRB*W9* lineage seems to be highly related to the *-DRB3*01/02* lineages. The present data elucidated that the above- mentioned lineages possess a highly similar microsatellite. Furthermore, an analysis of the introns of the *Patr-DRB3*01/02* and *-DRB*W9* alleles in the CENSOR database showed that intron 1 is similar in length and possesses the same Alu inserts (29). The similarity observed between the *Patr-DRB*W9* and *-DRB3*01/02* lineages must reflect the common ancestry. Thus, the alleles that were given the *Patr-DRB*W9* designations in the past are more likely alleles of a lineage of the *Patr-DRB3* locus. For a more consistent nomenclature, we renamed the *Patr-DRB*W9* alleles as follows: *Patr-DRB*W901* became *-DRB3*0701*, *-DRB*W902* became *-DRB3*0702*, and *-DRB*W903* became *-DRB3*0703*. The renaming is documented at the IPD-MHC-NHP-database (www.ebi.ac.uk/ipd/mhc/nhp).

Results and discussion

Patr-DRB haplotype definition by microsatellite analysis

In the past, restriction fragment length polymorphism (RFLP) followed by sequencing of the exon 2 segment of the *DRB* loci and Denaturing Gradient Gel Electrophoresis (DGGE) were used as techniques to describe the *Patr-DRB* profile of the West African chimpanzee colony, and resulted in the description of six different region configurations (6, 33, 34). The region configurations and haplotypes were firmly established based on segregation analyses. The present analyses, using the D6S2878 marker, showed that all exon 2- positive *Patr-DRB* genes possess the DRB-STR repeat. In the chimpanzee samples, 6–10 DRB-STR amplicons could be detected, which have highly variable lengths ranging from 135 to 226 base pair (bp). Subsequent sequencing of exon 2/DRB-STR alleles made it possible to link unequivocally a microsatellite to a particular exon 2. This approach resulted in a more accurate definition of *Patr-DRB* haplotypes (Table 1), and for the *P.t.v.* colony seven different region configurations can now be distinguished (Fig. 1A). The analyses of the *P.t.t.* and *P.t.s.* animals resulted in the detection of two additional region configurations (Fig. 1B), each composed of an exclusive combination of *Patr-DRB* genes: the combination of DRB-STR alleles seems to be unique for a given haplotype. Additionally, some DRB-STR alleles are highly predictive for the presence of a particular *DRB* allele. For example, the 157 fragment is observed in combination with *Patr-DRB1*0701*, the 155 fragment is observed in combination with *Patr-DRB6*0305*, and the 135 fragment segregates in combination with *Patr-DRB7*0101* (Table 1). There are also a few haplotypes in which the microsatellite displays length differences for a particular *Patr-DRB* allele: for instance, the *DRB1*0201*-linked DRB-STR of haplotype Ia (Table 1A). These slight length differences originate from different founder haplotypes, as was confirmed by segregation analysis.

In humans, five *DRB* region configurations are recognised and all display an abundant amount of allelic variation for the *DRB1* locus. Another situation is detected in rhesus and cynomolgus macaques. The latter species possess a high number of region configurations that display hardly any allelic variation within a given configuration (7, 11). The present data demonstrate that chimpanzees show a moderate amount of allelic variation within certain - but not all nine - described region configurations (Table 1). In general, the situation involving chimpanzees seems to resemble more closely the one found for the *HLA-DRB* region configurations. The human and chimpanzee region configurations contain similar *DRB* genes, illustrating that these genes must have been present in a common ancestor. For instance, region configuration IV is completely identical between humans and chimpanzees, indicating that it was present in a common ancestor and inherited 'untouched' during primate evolution. The stability of this region configuration may be due to an antisense integration of a retroviral insert within the *DRB7* pseudogene (35). Moreover, all *P.t.v.* region configurations are found to contain a *DRB1* gene, which is comparable to humans. In the *P.t.t.* and *P.t.s.* animals analysed region configuration IX was found to lack a *DRB1* gene, but most likely the prominent

Table 1. *Patr-DRB* haplotypes defined by DRB-STR genotyping and exon 2 sequencing for the *P.t.v.* subspecies (A) and *P.t.t.* (B) and *P.t.s.* subspecies (C). The haplotypes are arranged based on *DRB1* locus lineages; therefore, haplotype VII is situated between haplotype II and III. The last column specifies the number that each haplotype is present in the genotyped cohort. For alleles marked in yellow, a transcript has been found, and the full-length sequences are deposited at the EMBL database. Newly described exon 2 alleles are presented in bold. Data in parentheses indicate STR lengths detected only once or twice. No cell-line available for this haplotype (blue square). The alleles *Patr-DRB3*0701/0702/0703* were previously named *Patr-DRB*W901/W902/W903*, respectively (red square). Allele sequenced using *Patr-DRB1*02* specific primers (5'ACGTTTCCTGTTGCAGCC3' and 5'CCCCGTAGTTGTGTCTGC3') and expected to be linked to the STR allele 195 (green square). The STR length represents a sequencing result and therefore differs from the other *DRB6*0109* STR length (orange square).

	Haplotype	*DRB1* locus	STR	*DRB3* locus	STR	*DRB6* locus	STR	*DRB5* locus	STR	*DRB7* locus	STR	*DRB4* locus	STR	#
A	*Pan troglodytes verus*													
	Ia	*DRB1*0201*	207, 209, 211, 217	*DRB3*0201*	186	*DRB6*0108*	144	*DRB5*0301*	174					41
	Ib	*DRB1*0201*	209	*DRB3*0208*	220	*DRB6*0109*	145	*DRB5*0306*	174					20
	Ic	*DRB1*0201*	209	*DRB3*0208*	204	*DRB6*0108*	145	*DRB5*0304*	174					6
	Id	*DRB1*0201*	205	*DRB3*0208*	202	*DRB6*0108*	145	*DRB5*0307*	174					7
	Ie	*DRB1*0204*	205	*DRB3*0201*	188	*DRB6*0108*	144	*DRB5*0301*	174					6
	If	*DRB1*0204*	205, 221, 225	*DRB3*0208*	208, 212	*DRB6*0109*	145	*DRB5*0101*	170					35
	Ig	*DRB1*0204*	205	*DRB3*0208*	208 (200)	*DRB6*0109*	145	*DRB5*0102*	170					22
	Ih	*DRB1*0204*	205	*DRB3*0208*	210	*DRB6*0109*	145	*DRB5*0103*	170					1
	Ii	*DRB1*0214*	228	*DRB3*0103*	186	*DRB6*0109*	145	*DRB5*0101*	170					1
	IIa	*DRB1*0302*	182	*DRB3*0102*	186	*DRB6*0305*	155	*DRB5*0312*	174					6
	IIb	*DRB1*030701*	164, 174	*DRB3*0102*	186	*DRB6*0305*	155	*DRB5*0301*	174					5
	IIc	*DRB1*030702*	166	*DRB3*0102*	186	*DRB6*0305*	155	*DRB5*0306*	174					2
	VII	*DRB1*0305*	166, 168	*DRB3*0208*	214, 218, 222 (208)	*DRB6*0108/0305*	145/155	*DRB5*0310*	174					38
	IIIa	*DRB1*0302*	190 (184)	*DRB3*0102*	186	*DRB6*0305*	155							3
	IIIb	*DRB1*0309*	176	*DRB3*0102*	186	*DRB6*0305*	155							2
	IIIc ■	*DRB1*0309*	176	*DRB3*0703* ■	190	*DRB6*0305*	155							3
	IIId	*DRB1*0311*	184	*DRB3*0701* ■	190	*DRB6*0305*	155							1
	IVa	*DRB1*0701*	157							*DRB7*0101*	135	*DRB4*0104*	186	8
	V	*DRB1*1001*	170	*DRB3*0208*	208	*DRB6*0108*	145	*DRB5*0310*	174					2
	VI	*DRB1*1001*	168			*DRB6*0108*	145	*DRB5*0310*	174					4
B	*Pan troglodytes troglodytes*													
	Ij	*DRB1*0202*	226	*DRB3*0209*	206	*DRB6*0109*	145	*DRB5*0313*	195					2
	Ik ■	*DRB1*0215*	211	*DRB3*0208*	212	*DRB6*0109*	145	*DRB5*0311*	172					2
	IIIe	*DRB1*0308*	178	*DRB3*0102*	186	*DRB6*0305*	155							1
	IVb	*DRB1*0702*	153							*DRB7*0101*	135	*DRB4*0104*	186	3
	VIIIa	*DRB1*0205*	214	*DRB3*0214*	182									1
	IXa			*DRB3*0702* ■	192, 194							*DRB4*0201*	138	2
C	*Pan troglodytes schweinfurthii*													
	Il	*DRB1*0202* ■	195	*DRB3*0208*	212	*DRB6*0109*	143 ■	*DRB5*0311*	172					1
	VIIIb	*DRB1*0213*	213, 217	*DRB3*0214*	182									2
	IXb			*DRB3*0702* ■	192							*DRB4*0201*	138	1

task of the *HLA-DRB1* gene was taken over by one of the other genes on this region configuration. Chimpanzee region configuration I shows the highest degree of polymorphism and this is mostly due to the allelic variation within the *Patr-DRB5*03* lineage. Haplotype Ii is only observed in one animal, and contains an allele of the *Patr-DRB3*01* lineage, whereas all the other haplotypes of region configuration I possess an allele of the *Patr-DRB3*02* lineage (Table 1A). Nevertheless, a similar haplotype is present in the chimpanzee genome database (contig NM_01236523), thus confirming the existence of haplotype Ii. Another lineage that possesses several alleles is the *Patr-DRB1*03* lineage. It is the most polymorphic lineage, and distinct alleles are present on three different region configurations in the *P.t.v.* population studied (Table 1A). A similar observation has been made with regard to the *Patr-DRB3* locus, which is present on several different region configurations and shows lineage and allelic polymorphism.

The different *Patr-DRB* alleles were analysed phylogenetically together with *HLA-DRB* alleles from a Caucasoid population (Fig. 2A). The results equal the earlier published phylogenetic analysis on *DRB* exon 2 sequences (34), and demonstrate that chimpanzee and human alleles of identical lineages/loci cluster tightly together. In addition, the composition of the corresponding microsatellites is shown. The DRB-STR is a composite microsatellite, and the comparison of different *DRB* sequences of several primate species indicated that the ancient structure of the microsatellite most likely must have been (GT)x(GA)y (7, 10, 36, 37). Thus far, most *DRB* gene-associated microsatellites encountered are constructed of four sections: namely, a (GT)x, a (GA)z-mix, a (GA)y, and a (GC)n part. The 5'(GT)x repeat represents the longest segment, is mostly uninterrupted by other nucleotides, and is therefore known to evolve rapidly. For example, the *HLA-DRB5*010101* allele can be detected with different DRB-STR lengths, (GT)18-24 (Fig. 2B). This illustrates that the repeat itself evolves faster, due to a higher mutation rate, than the adjacent exon 2 sequence. Interruption of a repeat by other nucleotides usually results in more stability. This is, for instance, evidenced by the *Patr-DRB5* alleles, which all have a highly similar (GT)x part that is interrupted by a GA (Fig. 2B). The (GA)z-mix part is mostly shorter and interrupted by other dinucleotides, and appears to correlate with the different *DRB* lineages (Fig. 2B). Identical/similar color codes highlight that within particular lineages the relevant microsatellite is quite similar between humans and chimpanzees (Fig. 2B). Thus, not only the exon 2 sequences but also the DRB-STR alleles appear to reflect the common ancestry of *DRB* lineages. The (GA)y part often correlates with lineages as well, although this is less prominent. The (GC)n part forms the 3' end of the repeat. The *Patr-DRB1*1001* allele was found not to cluster together with its human equivalent, and both the *HLA-* and *Patr-DRB1*1001* alleles appear to possess a different microsatellite (Fig. 2). These observations suggest that the *HLA-* and *Patr-DRB1*1001* alleles belong to different lineages, and therefore the *Patr-DRB1*1001* allele is currently under investigation to unravel its evolutionary history. The presence of two different *Patr-DRB6* loci on one chromosome is unique for haplotype VII (Table 1A). The compositions of the microsatellites of the loci vary (Fig. 2B); indicating that most likely the haplotype arose

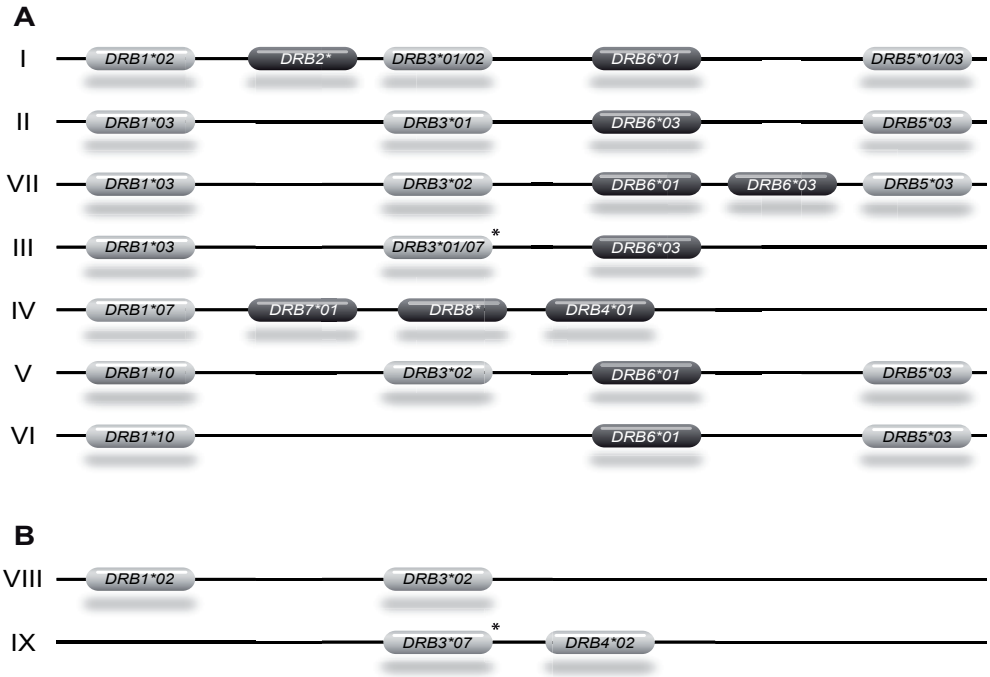
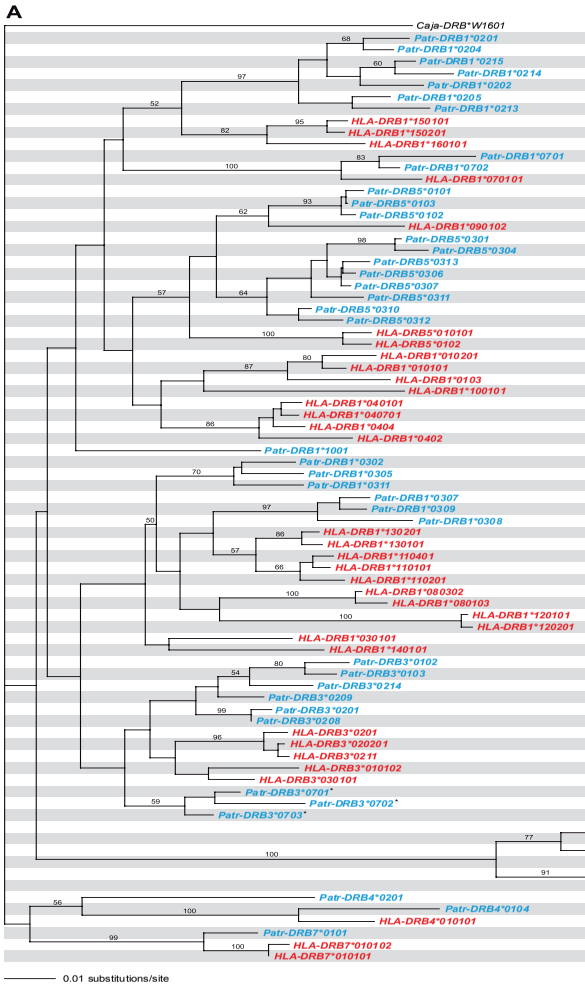


Fig. 1. A) *DRB* region configurations in the *P.t.v.* subspecies. The haplotypes are arranged based on *DRB1* locus lineages; therefore, haplotype VII is situated between haplotype II and III. B) Additional *DRB* region configurations detected in the *P.t.t.* and *P.t.s.* subspecies. The *DRB9* gene is not included but is present on all region configurations. Black boxes represent pseudogenes. For example, in configuration I *DRB3*01/02* means that lineage polymorphism is detected for this locus. Configuration IV is shared between humans and chimpanzees. *The *Patr-DRB3*07* lineage was previously named *Patr-DRB*W9* (details in material and methods).

via recombination rather then via duplication of one of the *DRB6* genes.

As compared to *Patr-DRB* typing strategies used earlier, the DRB-STR typing protocol appears to be an accurate and speedy method. The present analysis illustrates that next to humans and old world monkeys (7, 11), great ape species can also be haplotyped with the D6S2878 microsatellite marker. With the DRB-STR protocol it is possible to haplotype several different kinds of species for the *Mhc-DRB* region; hence, this microsatellite may find its application in pedigree analysis and can be used for conservation biology and breeding management in zoos.



B

Part 1 (GT)x	Part 2 (GA)z-mix	Part 3 (GA)y	Part 4 (GC)n
GTGA(GT)19,20,22	(GA)12-14CA(GA)4CA(GA)3GGAA	(GA)6,7	(GC)2
GTGA(GT)17,24,27	(GA)13,14CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
GTGA(GT)21,22	(GA)6,7CA(GA)5CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
GTGA(GT)25GA(GT)6	(GA)9CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
GTGA(GT)27,28	(GA)13,15CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
GTGA(GT)22	(GA)13CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
GTGA(GT)24	(GA)10CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
(GT)15-20	(GA)5,6CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
(GT)27-32	(GA)2CA(GA)4CA(GA)3GGAA	(GA)6	(GC)2
(GT)17	(GA)9GGAA	(GA)6	(GC)2
(GT)2GGTT(GT)14	(GA)4GC	(GA)2	CCGC
(GT)2GGTT(GT)10,11	(GA)7GC	(GA)2	CCGC
(GT)11	(GA)8GC	(GA)2	CCGC
(GT)4GA(GT)7	(GA)7GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GT)4GA(GT)7	(GA)7GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GT)4GA(GT)7	(GA)7GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GT)10,18	(GA)9,11(GG)1,2AA(GA)4CAGAAAGAGG	GA	(GC)3
(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)8	(GA)9GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)6	(GA)21GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)8	(GA)9,10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)6	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)4GA(GT)7	(GA)10GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)18-24	(GA)5-8GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)22	(GA)8GGAA(GA)4CA(GA)7GGAA(GA)4CA(GA)2GG	GA	(GC)3
(GT)19	AAGAAA	(GA)4	(GC)3
(GT)16,17	AAGAAA	(GA)4	(GC)3
(GT)16	AAGAAA	(GA)4	(GC)3
(CT)2(GT)16		(GA)8	(GC)3
(GT)21,22		(GA)15,16	(GC)2
(GT)22		(GA)17	(GC)2
(GT)22		(GA)21	(GC)2
(GT)23		(GA)18	(GC)2
(GT)9,10	(GA)11,12CA(GA)4AA	(GA)5	(GC)3
(GT)23,24,28	(GA)5AA	(GA)6	(GC)4
(GT)16-18	(GA)4,5AA	(GA)7	(GC)3
(GT)23,24	(GA)5AA	(GA)6	(GC)4
(GT)11-13		(GA)17,18,23	(GC)2
(GT)10,11	(GA)20,21CA	(GA)4	(GC)2
(GT)9	(GA)23CA	(GA)4	(GC)2
(GT)20,30,31	(GA)10-12CA(GA)3CA	(GA)6	(GC)2
(GT)22-24	(GA)10,11CA(GA)3CA	(GA)6	(GC)2
(GT)26	(GA)5CA(GA)3CA	(GA)5	(GC)2
(GT)22-26	(GA)5CA(GA)3CA	(GA)6	(GC)2
(GT)24	(GA)6CA(GA)3CA	(GA)6	(GC)2
(GT)21	(GA)7CA	(GA)5	(GC)3
(GT)18,20	(GA)7CA	(GA)5	(GC)3
(GT)25-29	(GA)11CA	(GA)10	(GC)2
(GT)32	(GA)10CA	(GA)10	(GC)2
(GT)17-22	(GA)6CA(GA)3CA	(GA)5	(GC)2
(GT)22-27	(GA)9-11CA	(GA)6	(GC)2
GA(GT)7AT(GT)5,6	(GA)5,6AA(GA)6GGAA(GA)4CA(GA)3GG	GA	(GC)3
GA(GT)8AT(GT)7	(GA)5AA(GA)6GGAA(GA)4CA(GA)3GG	GA	(GC)3
GA(GT)7AT(GT)5	(GA)6AA(GA)4GGAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)10,13(GT)15	(GA)10GGAA(GA)4CA(GA)3AG	GA	(GC)3
(GA)5(GT)13,14	(GA)9GGAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)4-9(GT)14-26	(GA)9-19(GG)2AA(GA)4CA(GA)3AG/GG	GA	(GC)3
GA(GT)22	(GA)15GGAA(GA)4CA(GA)3GG	GA	(GC)3
GA(GT)18-28	(GA)11-20GGAA(GA)4CA(GA)3GG	GA	(GC)3
GA(GT)23	(GA)16GGAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)3(GT)13,14	(GA)9,10GGAA(GA)2CA(GA)3GG	GA	(GC)3
GA(GT)7AT(GT)3AT(GT)5	(GA)6CA(GA)2GGAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)2(GT)7AT(GT)9	(GA)10GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)2(GT)7AT(GT)10	(GA)10GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GA)2(GT)7AT(GT)9	(GA)10GCAA(GA)4CA(GA)3GG	GA	(GC)3
(GT)4	GAGGGCAGGGGTCG(GG)3CAG	(GA)6	
(GT)4	GAGGGCAGGGGTCG(GG)3CAG	(GA)6	
GC(GT)3	GGGAGGAG(GG)3TGGA(GG)2CGATAGG	(GA)8	
(GT)15		(GA)6	
(GT)13-24		(GA)11-14	
GTCT(GT)6		(GA)10	
GA(GT)6GCGG	GGAA	(GA)5	GCACGC
GTAT(GT)4,5	(GA)8,9(CAGA)2(GA)10CAGATGAA(GA)3AA	GA	GCGT(GC)2
GTAT(GT)9-11	(GA)9-14(CAGA)1,2GGAA	(GA)5	GC(GT)1,2(GC)1,2
(GT)2TT(GT)3T	(GA)6(CA)5		
(GT)2TT(GT)3T	(GA)5(CA)6		
(GT)2TT(GT)3T	(GA)5(CA)6		

Fig. 2. (A) Rooted phylogenetic comparison of the exon 2 sequences of *DRB* alleles detected in chimpanzees (*Patr*, blue) and humans (*HLA*, red). The common marmoset allele *Caja-DRB*W1601* is taken as an outgroup (48). Relevant bootstrap values are indicated. The brackets specify the different *DRB* loci/lineages. *The *Patr-DRB3*07* lineage was previously named *Patr-DRB*W9*. (B) Composition of the microsatellite sequence (D6S2878) corresponding to the *DRB* allele in (A). Identical or similar colours indicate identity or similarity of the (GA)z-mix part of the microsatellite.

Full-length cDNA analysis of Patr-DRB alleles

To determine which *Patr-DRB* alleles are transcribed, a panel of cell lines of different animals was selected that covered most of the known haplotypes (Table 1A/B). All *Patr-DRB1* and *-DRB5* alleles present on the different haplotypes were found to produce bona fide transcripts. For the *Patr-DRB3* locus, we detected transcripts for the *-DRB3*0102*, *-DRB3*0103*, *-DRB3*0201*, *-DRB3*0214*, and *-DRB3*07* alleles. For the *Patr-DRB3*0208* and *-DRB3*0209* alleles, no transcripts were observed. The *Patr-DRB3*0309* is observed only on one haplotype and for only one animal a cell-line was available to analyse this allele. Thus, there is a possibility that due to primer inconsistencies the allele was not amplified, and therefore this allele needs further investigation. The *Patr-DRB3*0208* allele appears to be a pseudogene, as sequencing analysis on genomic DNA resulted in the detection of a premature stopcodon in exon 3 (38). No transcripts were detected for the *Patr-DRB4*0104* and *-DRB7*0101* alleles, present on region configuration IVa/b (Table 1A/B). *Patr-DRB4*0104* was deemed to represent a non-functional allele because of the presence of a premature stopcodon at the end of exon 2, whereas the *-DRB7*0101* allele is considered to be a pseudogene because of the presence of deletions and a premature stopcodon (34, 35). However, the *Patr-DRB4*0201* allele, representing the other lineage within the *Patr-DRB4* locus, produces a transcript. Previously, it was described that this allele most likely arose from a recombination event between a *Patr-DRB4*01* allele and an unknown donor allele (34). The data indeed illustrate that the composition of the DRB-STR is different between the alleles of the *HLA-* and *Patr-DRB4*01* and *Patr-DRB4*02* lineages (Fig. 2B). The comparison of the full-length sequences of the *HLA-DRB4*0101* and *Patr-DRB4*0201* alleles shows that the sequences diverged, and only a small motif in exon 2 characteristic for the *DRB4* locus is shared between the alleles. These observations suggest a hybrid character for the *Patr-DRB4*0201* allele, and therefore currently the introns of this allele are sequenced to unravel its evolutionary history. Moreover, with the primer set used, no full-length transcripts are observed for the *DRB6* locus. This is expected, as this locus lacks its exon 1 and is therefore considered to be a pseudogene (34, 39). Taken together, the present communication increases the number of known full-length *Patr-DRB* sequences from 3 to 32. The full-length sequences of the *Patr-DRB1*0201*, *-DRB3*0201*, and *-DRB5*0301* alleles are identical to the earlier published sequences W1, C4-2, and B3-5, respectively (13).

The *Patr-DRB5*0313* allele (present on haplotype Ij) detected in one of the *P.t.t.* animals is very similar to *Patr-DRB5*0306* (present on haplotype Ib and IIc). Although, at position 13 in exon 2 of the *Patr-DRB5*0313* allele, the triplet TGT coding for a cysteine is substituted by CAT, which encodes for a histidine. A histidine is rare at this position and is only observed in alleles of the *HLA-DRB1*04* and *Patr-DRB4*01* lineages (Fig. 3). However, chimpanzees are known to lack the equivalent of the *HLA-DRB1*04* lineage, and *Patr-DRB4*01* seems to represent a non-functional lineage. As such, the substitution of the cysteine by a histidine in the *Patr-DRB5*0313* allele suggests the rescue of a residue, most probably by recombination, originating from a different functional chimpanzee lineage.

```
                    1       10        20
HLA-DRB1*010101     GDTRPRFLWQLKFECHFFNG
HLA-DRB1*040101     --------E-V-H-------
Patr-DRB1*0201      ******---L-P-G-------
Patr-DRB1*0302      ******---EYSTS-------
Patr-DRB1*0701      ******------S-YK------
Patr-DRB1*1001      ******---E-A-C-------
Patr-DRB3*0102      ******---ELH-S-------
Patr-DRB3*0201      ******---ELV-S-------
Patr-DRB4*0104N     ******---E-A-H----L--
Patr-DRB4*0105      ******---E-A-H----L--
Patr-DRB4*0106      *******-E-A-H----L--
Patr-DRB4*0107      *******-E-A-H----L--
Patr-DRB4*0201      *******-E-----------
Patr-DRB5*0101      ******---K-D-Y-------
Patr-DRB5*0301      ******---K-D-Y-------
Patr-DRB5*0306      --------K-D-C-------
Patr-DRB5*0313      --------K-D-H-------
Patr-DRB6*0105N     ******---EKA-C---I---
Patr-DRB6*0301N     ********E-A-C---I---
Patr-DRB7*0101N     ******---E-A-S-------
```

Fig. 3. Deduced protein alignment of the beginning of exon 2 of a representative allele of the *HLA-DRB1*04* lineage and representative alleles of the different *Patr-DRB* lineages. The *HLA-DRB1*010101* is taken as consensus from the IMGT/HLA database (www.ebi.ac.uk/imgt/hla). Identity to the consensus sequence is indicated by dashes, whereas amino acid replacements are depicted by the conventional one letter code. The *Patr-DRB5*0306* and *-DRB5*0313* alleles are marked in orange. The histidine at position 13 is marked in green. N means null-allele.

In humans, only haplotypes with one or two transcribed *DRB* alleles are known (32). Most chimpanzee haplotypes of the *P.t.v.* subspecies possess two transcribed *DRB* alleles. Only haplotype IVa contains one, whereas haplotypes Ia, Ie, Ii, and IIa to IIc contain three *DRB* alleles that are potentially translated into a polypeptide (Table 1A). The analysed animals of the *P.t.t.* subspecies possess haplotypes with one (r.c. IVb) or two (r.c. IIIe, VIIIa, and IXa) transcribed *DRB* allele(s) (Table 1B). In comparison, rhesus macaques, which possess haplotypes with up to six different *Mamu-DRB* genes, are known to transcribe one to three *DRB* genes/alleles (40). In humans, the presence of only one transcribed *DRB* gene per haplotype (HLA-DR1 and -DR8 serotypes) does not seem to have any disadvantage, as HLA-DR1 and DR8 homozygous individuals are not rare and appear to be healthy (32, 41). Nonetheless, the transcription of three *DRB* alleles per haplotype is most probably the limit, as more *DRB* transcripts per haplotype have not yet been reported. This assumption seems to be plausible if one envisages that the expression of too many MHC class II molecules would result in a high number of T cells that would be deleted during thymic education. As a consequence, such individuals would possess a reduced T-cell repertoire and therefore may become susceptible to infectious diseases.

The complete cDNA sequences of the analysed *Patr-DRB* alleles and particular *HLA-DRB* alleles were subjected to phylogenetic analysis (Fig. 4). The analysis confirms the results of the exon 2 phylogenetic comparison (Fig. 2A and (34), and illustrates that for most of the full-length *HLA-* and *Patr-DRB* genes a trans-species mode of evolution is observed (42). Even if intron sequences of both species are compared, the footprint of a trans-species mode of evolution is retained. This is, however, different from the situation observed in rhesus macaques. The *Mamu-DRB* exon 2 sequences represent old entities that predate primate speciation, whereas these exons are embedded in genes that appear to represent relatively young entities (38). The full-length sequence of *Patr-DRB3*0103* appears to cluster with *Patr-DRB3*0201* (Fig. 4). This is different from what observed in the exon 2 phylogenetic comparison (Fig. 2A), and indicates that also in chimpanzees exon 2 shuffling in the *Patr-DRB* genes may appear. This seems to be, however, a rare event.

Chimpanzee DRB alleles and their potential contribution to resistance to AIDS

Analysis of the MHC region revealed that chimpanzees experienced a selective sweep affecting mainly the *Mhc* class I repertoire. This selective sweep may have been caused by a HIV-1/SIV_{cpz}-like retrovirus (20). In addition to *Mhc* class I, other parts within the *Mhc* region also show signs of a repertoire reduction, such as the major histocompatibility chain-related gene (*MIC*), located centromeric of the *Mhc-B* locus (21, 43). Comparative genomics showed that the strongest repertoire reduction maps to the chimpanzee *Mhc* class I region. However, different parts in the *Mhc* class II region show signs of a repertoire reduction as well (6). A recent study performed in a cohort of Kenyan sex workers reported that particular *HLA-DRB* alleles are associated with resistance or susceptibility to HIV-1 mediated disease (44). We compared these alleles with the *Patr-DRB* alleles present in our cohort.

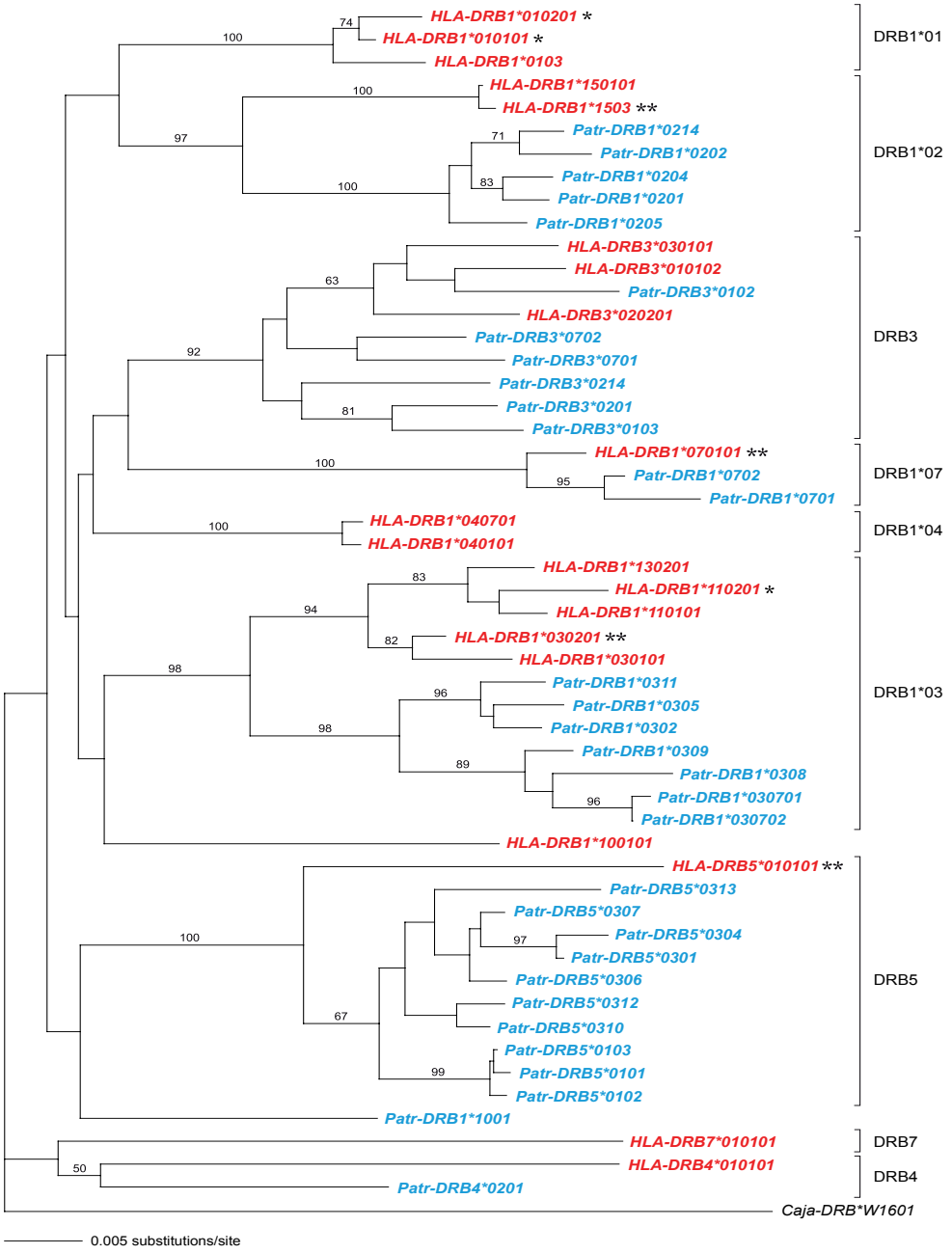


Fig. 4. Rooted phylogenetic comparison of the full-length chimpanzee (*Patr*, blue) cDNA sequences. Particular *HLA-DRB* alleles (red) extracted from the IMGT/HLA-database were included (32), the sequences of *HLA-DRB1*110201* and *-DRB1*030201* were not full-length, but this did not alter the phylogenetic comparison. The *Caja-DRB*W1601* allele is taken as an outgroup (48). Relevant bootstrap values are indicated. The brackets specify the different *DRB* loci/lineages. A * indicate the *HLA-DRB* alleles associated with resistance and ** indicate the alleles associated with susceptibility to HIV-1 infection found in a cohort of Kenyan sex workers.

Equivalents of the *HLA-DRB1*070101*, *-DRB1*1503*, and *-DRB5*010101*, associated with susceptibility, can be found in chimpanzees (Fig. 4). For the *HLA-DRB1*010101* and *010201* alleles, associated with resistance, no apparent equivalents are detected in chimpanzees. The *HLA-DRB1*1102* allele, associated with resistance, seem to cluster phylogenetically together with alleles of the *Patr-DRB1*03* lineage. However, the *HLA-DRB1*030201* allele, associated with susceptibility, seems to cluster as well with the *Patr-DRB1*03* lineage. This suggests that it is difficult to define a contribution of the *Patr-DRB* lineages in HIV-1/SIV_{cpz} infection based on phylogenetic comparisons of sequences. Indeed, in humans it has recently been shown that each of the nine antigens of the HLA-B44 supertype possess unique peptide binding motifs (45). This observation, together with the knowledge that the peptide binding affinity of a MHC class II molecule is less stringent than the binding affinity of a MHC class I molecule, corroborates the notion that caution must be exercised in drawing conclusions based only on the similarity observed in phylogenetic comparisons (46, 47). With the availability of the full-length cDNAs of different *Patr-DRB* alleles, we are now able to construct single antigen-expressing cell lines and study in detail the function and the MHC/peptide interaction of the Patr-DRB molecules.

Nevertheless, the human study indicated that the DRB-specific $CD4^+$ T-cell responses are an important factor in resistance/susceptibility to HIV-1 infection (44). For chimpanzees, it is known that they maintain $CD4^+$ T-cell responses after HIV-1/SIV_{cpz} infection. Comparing the *HLA-* and *Patr-DRB* region, it is obvious that both species have a slightly different number of region configurations, but the region configurations possess identical *DRB* genes. Humans on the one hand seem to have acquired a large amount of allelic variation within the different *DRB* region configurations, whereas chimpanzees on the other hand, although known to be an older species, show only limited allelic variation. Additionally, chimpanzees lack the equivalent of the *HLA-DRB1*04* lineage. These observations support that the chimpanzee MHC class II region may have experienced a reduction in the repertoire as well. The contemporary MHC class II alleles were most likely positively selected due to genetic linkage to the MHC class I alleles that survived the selective sweep. The presently available repertoire of *Patr-DRB* alleles seems to contribute efficiently to the maintenance of the $CD4^+$ T-cell responses, and therefore chimpanzees seem to have successful combinations of $CD4^+$/$CD8^+$ T-cell responses that are able to cope with pathogens of their natural habitat and with HIV-1/SIV_{cpz} infection.

Acknowledgements

The authors wish to thank D. Devine for editing the manuscript, H. van Westbroek for preparing the figures, and Drs F. Claas and I. Doxiadis for providing the human DNA samples. This study was supported in part by NIH/NIAID contract number HHSN266200400088C and 5R24RR016038-05.

References

1. Fujiyama A, Watanabe H, Toyoda A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
2. Gonder MK, Oates JF, Disotell TR, et al. (1997) A new west African chimpanzee subspecies? Nature 388:337.
3. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
4. Marsh SG, Albert ED, Bodmer WF, et al. (2005) Nomenclature for Factors of the HLA System, 2004. Hum Immunol 66:571-636.
5. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
6. de Groot NG, Heijmans CM, de Groot N, et al. (2008) Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. Mol Ecol 17:2074-2088.
7. Doxiadis GG, de Groot N, Claas FH, et al. (2007) A highly divergent microsatellite facilitating fast and accurate *DRB* haplotyping in humans and rhesus macaques. Proc Natl Acad Sci U S A 104:8907-8912.
8. Andersson G, Larhammar D, Widmark E, et al. (1987) Class II genes of the human major histocompatibility complex. Organization and evolutionary relationship of the DR beta genes. J Biol Chem 262:8748-8758.
9. Epplen C, Santos EJ, Guerreiro JF, et al. (1997) Coding versus intron variability: extremely polymorphic HLA-DRB1 exons are flanked by specific composite microsatellites, even in distant populations. Hum Genet 99:399-406.
10. Riess O, Kammerbauer C, Roewer L, et al. (1990) Hypervariability of intronic simple (gt)n(ga)m repeats in *HLA-DRB* genes. Immunogenetics 32:110-116.
11. de Groot N, Doxiadis GG, de Vos-Rouweler AJ, et al. (2008) Comparative genetics of a highly divergent DRB microsatellite in different macaque species. Immunogenetics 60:737-748.
12. Bak EJ, Ishii Y, Omatsu T, et al. (2006) Identification and analysis of MHC class II DRB1 (*Patr-DRB1*) alleles in chimpanzees. Tissue Antigens 67:134-142.
13. Fan WM, Kasahara M, Gutknecht J, et al. (1989) Shared class II MHC polymorphisms between humans and chimpanzees. Hum Immunol 26:107-121.
14. Balla-Jhagjhoorsingh SS, Verschoor EJ, de Groot N, et al. (2003) Specific nature of cellular immune responses elicited by chimpanzees against HIV-1. Hum Immunol 64:681-688.
15. Heeney JL, Rutjens E, Verschoor EJ, et al. (2006) Transmission of simian immunodeficiency virus SIVcpz and the evolution of infection in the presence and absence of concurrent human immunodeficiency virus type 1 infection in chimpanzees. J Virol 80:7208-7218.
16. Keele BF, Jones JH, Terio KA, et al. (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460:515-519.
17. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
18. Goulder PJ, Watkins DI (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol 8:619-630.
19. Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, et al. (1999) Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol 162: 2308-2314.

20. de Groot NG, Otting N, Doxiadis GG, et al. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci U S A 99: 11748-11753.
21. de Groot NG, Garcia CA, Verschoor EJ, et al. (2005) Reduced *MIC* gene repertoire variation in West African chimpanzees as compared to humans. Mol Biol Evol 22:1375-1385.
22. Williams MA, Holmes BJ, Sun JC, Bevan MJ (2006) Developing and maintaining protective $CD8^+$ memory T cells. Immunol Rev 211:146-153.
23. Bottius E, BenMohamed L, Brahimi K, et al. (1996) A novel Plasmodium falciparum sporozoite and liver stage antigen (SALSA) defines major B, T helper, and CTL epitopes. J Immunol 156:2874-2884.
24. Erickson AL, Houghton M, Choo QL, et al. (1993) Hepatitis C virus-specific CTL responses in the liver of chimpanzees with acute and chronic hepatitis C. J Immunol 151:4189-4199.
25. Bontrop RE, Otting N, Slierendregt BL, Lanchbury JS (1995) Evolution of major histocompatibility complex polymorphisms and T-cell receptor diversity in primates. Immunol Rev 143:33-62.
26. de Groot NG, Otting N, Arguello R, et al. (2000) Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. Immunogenetics 51:398-409.
27. Lekutis C, Letvin NL (1995) Biochemical and molecular characterization of rhesus monkey major histocompatibility complex class II DR. Hum Immunol 43:72-80.
28. Swofford D (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
29. Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7:474.
30. Ellis SA, Bontrop RE, Antczak DF, et al. (2006) ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics 57:953-958.
31. Klein J, Bontrop RE, Dawkins RL, et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217-219.
32. Robinson J, Waller MJ, Parham P, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31: 311-314.
33. Bontrop RE, Broos LA, Pham K, et al. (1990) The chimpanzee major histocompatibility complex class II DR subregion contains an unexpectedly high number of beta-chain genes. Immunogenetics 32:272-280.
34. Kenter M, Otting N, Anholts J, et al. (1992) *Mhc-DRB* diversity of the chimpanzee (*Pan troglodytes*). Immunogenetics 37:1-11.
35. Doxiadis GG, de Groot N, Bontrop RE (2008) Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. J Virol 82:6667-6677.
36. Bergstrom TF, Engkvist H, Erlandsson R, et al. (1999) Tracing the origin of HLA-DRB1 alleles by microsatellite polymorphism. Am J Hum Genet 64:1709-1718.
37. Kriener K, O'HUigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics 51:169-178.
38. Doxiadis GG, de Groot N, de Groot NG, et al. (2008) Reshuffling of ancient peptide binding motifs between *HLA-DRB* multigene family members: old wine served in new skins. Mol Immunol 45:2743-2751.

39. Mayer WE, O'HUigin C, Klein J (1993) Resolution of the HLA-DRB6 puzzle: a case of grafting a de novo-generated exon on an existing gene. Proc Natl Acad Sci U S A 90:10720-10724.
40. de Groot N, Doxiadis GG, De Groot NG, et al. (2004) Genetic makeup of the *DR* region in rhesus macaques: gene content, transcripts, and pseudogenes. J Immunol 172:6152-6157.
41. Marsh SG, Parham, P., Barber, L.D. (2000) The HLA FactsBook, Academic Press, London, UK.
42. von Salome J, Gyllensten U, Bergstrom TF (2007) Full-length sequence analysis of the HLA-DRB1 locus suggests a recent origin of alleles. Immunogenetics 59:261-271.
43. Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122.
44. Lacap PA, Huntington JD, Luo M, et al. (2008) Associations of human leukocyte antigen DRB with resistance or susceptibility to HIV-1 infection in the Pumwani Sex Worker Cohort. Aids 22:1029-1038.
45. Hillen N, Mester G, Lemmel C, et al. (2008) Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. Eur J Immunol 38:2993-3003.
46. Hammer J, Valsasnini P, Tolba K, et al. (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. Cell 74:197-203.
47. Rammensee HG, Falk K, Rotzschke O (1993) MHC molecules as peptide receptors. Curr Opin Immunol 5:35-44.
48. Doxiadis GG, van der Wiel MK, Brok HP, et al. (2006) Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. Proc Natl Acad Sci U S A 103:5864-5868.

Chapter 7

AIDS-protective HLA-B*27/B*57 and chimpanzee MHC class I molecules target analogous conserved areas of HIV-1/SIV$_{cpz}$

Natasja G. de Groot[a], Corrine M.C. Heijmans[a], Yvonne M. Zoet[b], Arnoud H. de Ru[b], Frank A. Verreck[a], Peter A. van Veelen[b], Jan W. Drijfhout[b], Gaby G.M. Doxiadis[a], Edmond J. Remarque[a], Ilias I.N. Doxiadis[b], Jon J. van Rood[b], Frits Koning[b], and Ronald E. Bontrop[a,c]

Proceedings of the National Academy of Sciences USA 2010, in press

[a]Comparative Genetics and Refinement, and Parasitology, Biomedical Primate Research Centre, 2288 GJ Rijswijk, The Netherlands.
[b]Immunohaematology and Blood Transfusion, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.
[c]Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands.

Abstract

In the absence of treatment, most HIV-1-infected humans develop AIDS. However, a minority are long-term non-progressors, and resistance is associated with the presence of particular HLA-B*27/B*57 molecules. In contrast, most HIV-1-infected chimpanzees do not contract AIDS. In comparison to humans, chimpanzees experienced an ancient selective sweep affecting the MHC class I repertoire. We have determined the peptide binding properties of frequent chimpanzee MHC class I molecules, and show that like HLA-B*27/B*57 they target similar conserved areas of HIV-1/SIV_{cpz}. In addition, many animals appear to possess multiple molecules targeting various conserved areas of the HIV-1/SIV_{cpz} Gag protein, a quantitative aspect of the immune response that may further minimize the chance of viral escape. The functional charateristics of the contemporary chimpanzee MHC repertoire suggests that the selective sweep was caused by a lentiviral pandemic.

Introduction

In evolutionary terms, chimpanzees and humans are each other's closest living relatives (1), and both species are susceptible to infection with human immunodeficiency virus type 1 (HIV-1), which in particular individuals can lead to acquired immunodeficiency syndrome (AIDS). The HIV-1 pandemic in humans arose from multiple zoonotic events caused by the transmission of the chimpanzee-derived simian immunodeficiency virus (SIV_{cpz}) (2). Natural infections with various SIV_{cpz} strains have been documented in Central- and East-African chimpanzee populations (3). Recently, AIDS-like immunopathology was reported for some animals of the East-African sub-species (4). However, naturally infected chimpanzees of the same sub-species that have been monitored carefully during captivity over long periods of time – at our center – do not develop any signs of AIDS, despite relatively high viral plasma loads (5). Moreover, AIDS-like disease did not emerge in West-African chimpanzees after the experimental introduction of a natural SIV_{cpz} strain (5). Hence, chimpanzees are susceptible to HIV-1/SIV_{cpz} infection but are relatively resistant to progress towards AIDS (6, 7). On the contrary, most human subjects develop AIDS upon HIV-1 infection. However, cohort studies have established the existence of individuals that have been infected for more than 17 years but have not progressed to AIDS; these are the long-term non-progressors/elite controllers. Resistance has been attributed to infections with defective viruses or to mutations in the chemokine (C-C motif) receptor 5 of the host, the primary co-receptor for viral entry (8, 9). Furthermore, cytotoxic T-cell (CTL) responses play a prominent role in controlling intracellular viral infections. The T-cell receptor on such cells may recognize major histocompatibility complex (MHC) class I molecules complexed with a viral peptide, and subsequently the CTL will lyse the infected cell. In humans and chimpanzees, these molecules are designated HLA- and Patr-A, -B, and -C respectively, and they display abundant levels of polymorphism (10).

Resistance to progression to AIDS in human subjects is strongly associated with the presence of particular HLA-B*27/B*57 molecules, in conjunction with specific killer cell inhibitory receptors (11, 12). Genome-wide association studies have further substantiated the prominent role of the MHC class I region in controlling HIV-1 replication (13, 14).
We have reported several lines of genetic evidence that chimpanzees experienced a selective sweep that targeted the *Patr* class I repertoire, and hypothesized that this was caused by HIV-1/SIV$_{cpz}$ or a related ancestral retrovirus (15-17). The sweep, dated to approximately 2-3 million years ago, also matches the time frame with respect to when the chimpanzee genome might have been exposed to retroviral integrations (18). We wished to investigate whether the selective sweep was caused by HIV-1/SIV$_{cpz}$ or a closely related ancestor, and whether this repertoire skewing resulted in the preferential selection of Patr molecules that are similar to AIDS-resistant molecules in human individuals, such as HLA-B*27/B*57. For this purpose, we have determined the peptide binding motifs of Patr molecules that occur at a high frequency in a wild-caught chimpanzee population. Based on this information, the HIV-1/SIV$_{cpz}$ proteomes were scanned for potential CTL epitopes, which were subsequently tested in peptide binding studies. Attention was focused on the Gag (p24) protein as a modelsystem, as adaptive immune responses to this viral protein are thought to play a prominent role in the control of viral replication (19-21). In addition, cellular immune response data were re-evaluated in the context of the information gained from the peptide binding studies. The results demonstrate that most chimpanzees from the population investigated possess at least one Patr molecule that can bind analogous conserved areas of the HIV-1/SIV$_{cpz}$ Gag protein, as does HLA-B*27/B*57.

Results

Definition of peptide binding motifs

The peptide-binding site of MHC class I molecules generally accommodates nonamer peptides. The second residue of the peptide (numbered from the N-terminus) is often fixed as an anchor in the B-pocket, whereas the carboxyl-terminus is bound to the F-pocket (22). HLA- and Patr-class I gene products display polymorphism, and, as a consequence, different molecules possess distinct peptide binding motifs (10). The Patr-A and -B molecules present in the population studied were clustered based on the sharing of similar B- and F-pockets (Fig. 1). The molecules with the highest frequency within a given cluster (Patr-A*0301, -B*0101, -B*0301, and -B*0501) were selected for the determination of their peptide binding motifs. For this purpose, single molecule-expressing cell lines were constructed. Subsequently, the Patr-molecules were isolated by immunopurification, followed by extraction of the pool of natural bound peptides. We then identified the amino acid sequence of the isolated peptides by tandem mass spectrometry. Binding motifs were assembled from the isolated 8, 9, and 10-mer peptides (Table S1, Table S2 and Fig. S1).

MHC molecule	B-Pocket	F-Pocket	All. fr.	Peptide binding motif [a]
HLA-B*5701	**YYAVMENMSY**	**TNIAYSYYTKW**		**x-[ATS]-x-x-x-x-x-x-[FW]**
Patr-B*0901	--T-------	-----Y-----	1.8	x-[ST]-x-x-x-x-x-x-[WFLY]
Patr-B*0101	--T-------	-----Y-----	33.3	x-[ST]-x-x-x-x-x-x-[IL]
Patr-A*0301	----------	-DTL-D-----	17.5	x-[ST]-x-x-x-x-x-x-[RK]
Patr-A*0302	----------	-DTL-D-----	1.8	
Patr-A*1101	----------	--TL-D-----	1.8	
Patr-B*1401	-----NGIQ-	--TL-Y-----	8.8	
Patr-B*2901	-----NGVQ-	-SNL-Y-----	3.5	
Patr-B*0401	--T-K-ISNF	-GNL-Y-----	1.8	
Patr-B*16011	--T-K-ISNF	-SNL-Y-----	3.5	
Patr-B*2402[b]	--T-K-ISNF	--TL-Y-----	12.3	x-[DE]-x-x-x-x-x-x-[VIL(A)]
Patr-B*3001	--T-K-ISN-	-----Y-----		
HLA-B*2705	**-HT-E-ICK-**	**-DTL-D-----**		**x-[R(K)]-x-x-x-x-x-x-[LFYRHK(MI)]**
Patr-B*0301	-DT----V-F	-----Y-----	3.5	x-[R]-x-x-x-x-(L)-x-[IL]
Patr-B*1301	--S-ENIYQ-	-SNL-Y-----	12.3	x-[P]-x-x-x-x-x-x-[FLMIV]
Patr-B*0501	--S-E-ISN-	-SNL-Y-----	15.8	x-[KQ]-x-x-x-x-x-x-[L]
Patr-B*1101	--S-ENIYQ-	-GNL-Y-----		
Patr-B*1201	****ENIYQ-	-SNL-D-----		
Patr-B*1202	--S-ENICQ-	-SNL-D-----		
Patr-B*2303	--S-E-ISN-	--TL-Y-----		
Patr-B*2801	--T-ENIFQ-	-GNL-H-----		
Patr-A*0501	-S----SVFF	-DTL-D-----	7.0	
Patr-A*0401	-S----SV-F	-DTL-D-----	8.8	#-x-x-x-#-x-x-x-[RK]
Patr-A*0601	-S----SV--	--TL-D-----	7.0	
Patr-A*0602	-S----SVF-	--TL-D-----	1.8	x-x-x-x-x-x-x-x-[Y]
Patr-A*0901	-S----SV--	-DTL-Y-----	14.0	x-[WYF(AMVISNQ)]-x-x-x-x-x-x-[MLIVA(RFTS)]
Patr-A*1401[c]	-S----SVFF	--TL-H----C	7.0	
Patr-A*0701[d]	-S----SVGF	I-TL-L-----	12.3	x-[YFMLI(P)]-x-x-x-x-x-x-[VLMAFI]
Patr-A*1001	-S------FF	I-TL-H-----		
Patr-A*1201	-S----SV-F	--TL-H-----		
Patr-A*1301	-S----SVFF	--TL-Y-----		
Patr-A*0101	-F----SAH-	-DTL-D-----	10.5	x-[FMLI(VAST)]-x-x-x-x-x-x-[RK]
Patr-A*0201	-F----SAH-	-DTL-D-----	1.8	
Patr-A*1702	------SA--	-DTL-D-----		
Patr-B*2701	--T----V--	-SNL-Y-----		
Patr-B*2601	--T----VF-	--TL-F-----		
Patr-B*2501	-L------F-	-----F-----		

Fig. 1. Chimpanzee class I molecules clustered based on B- and F-pocket similarities. The B- and F-pocket of HLA-B*5701 are taken as consensus. A dash indicates identity to the consensus, an amino acid replacement is represented by the conventional one-letter code, and an asterisk indicates the position where the amino acid is unknown. The allele frequency (All. fr.) figures of the wild-caught founder population have been provided. Molecules depicted in light and dark brown are from chimpanzees of the Central- and East-African subspecies, respectively. Dark-orange represents unreported peptide binding motifs, whereas light orange highlights the motifs that are in agreement with data reported by another research team (23, 24). The binding motifs for HLA-B*2705 and -B*5701 as well as previously identified Patr binding motifs are provided (23-25, 43). [a]The conventional one-letter code identifies the preferred, or in brackets the tolerated, residues at the corresponding positions. [b]Patr-B*2402 has an identical B- and a similar F-pocket (--TL-F-----) to Patr-B*2401. Binding motif shown is from Patr-B*2401. [c]Molecule present in West- and East-African subspecies. [d]Molecule present in West-, Central-, and East-African subspecies. The # marks additional primary anchors.

The motifs elucidated for Patr-A*0301 and -B*0101 share their B-pocket anchor with HLA-B*5701 (Fig. 1), in accordance with data generated by others (23, 24). For Patr-A*0301, the F-pocket anchor is similar to HLA-A*03/11, whereas Patr-B*0101 prefers hydrophobic amino acids at this position, a phenomenon shared with different HLA molecules (25). The as yet unreported motifs for Patr-B*0301 and -B*0501 were either similar to those described for particular HLA-B*27 molecules or were identical to those of HLA-B*3902/B*4801, respectively (25). With the current data, the total number of peptide binding motifs defined for Patr class I molecules is twelve. This further broadens our ability to understand and study the MHC-peptide interactions in chimpanzees.

Prediction and selection of potential HIV-1/SIV$_{cpz}$ Gag epitopes

The peptide binding motifs allowed us to scan the entire Gag protein for potential epitopes using a database (26). We paid particular attention to those areas of Gag that are also recognized by HLA-B*27/B*57, as these epitopes may be crucial for the control of viral replication in humans (19, 27). These CTL epitopes have been designated IW9, KF11, TW10, KK10, and QW9 (Fig. 2). Based on the binding motifs, Patr-A*0301 and -B*0101 are predicted to recognize two peptides that overlap with the IW9 epitope (Fig. 2). In addition, Patr-B*0101 is predicted to bind a TW10-related epitope as well. The scan for Patr-B*0301 revealed that it can target overlapping areas of IW9 and KK10, whereas Patr-B*0501 can recognize different areas such as IW9, KF11, KK10, and QW9 (Fig. 2). Taken together, 10 different Gag peptides were selected for further analysis of their capacity to act as ligands for the four Patr molecules (Fig. 2, and Fig. S2).

Subsequently, an additional algorithm was applied as a tool to verify the binding affinity of the selected peptides (28, 29). Moreover, this algorithm was used to search for other potential CTL epitopes. Two peptides, HK9 and RL9, were chosen, as they showed high binding level scores for various Patr molecules (Fig. 2). In humans, CTL responses directed against HK9 and RL9 have been observed in conjunction with the restriction elements HLA-A*11 and -B*52 or -A*02, respectively (www.hiv.lanl.gov). The Gag-derived peptide NPPIPVGEI and a Gag-unrelated peptide EEALQAFTY, both predicted to have no or a very low binding affinity for each of the four molecules, were selected as controls.

Peptide binding studies

To determine the binding affinity of the selected Gag peptides, a cell-based peptide binding competition assay was developed. Each Patr molecule has its specific biotine-labeled indicator peptide against which the binding affinity of the different selected peptides was determined. The binding affinity of the indicator peptide itself was verified using its unlabeled peptide as competitor, henceforth referred to as standardized competitor (SC). Dose-inhibition curves for the peptides predicted to bind to Patr-A*0301, -B*0101, B*0301, and -B*0501 were derived from pooled data from five individual experiments (Fig. S3), and all showed a high or intermediate affinity to the relevant Patr molecules (Fig. 3, bold). The intrinsic binding affinity of the SC for Patr-B*0301 was high in comparison to the binding affinity

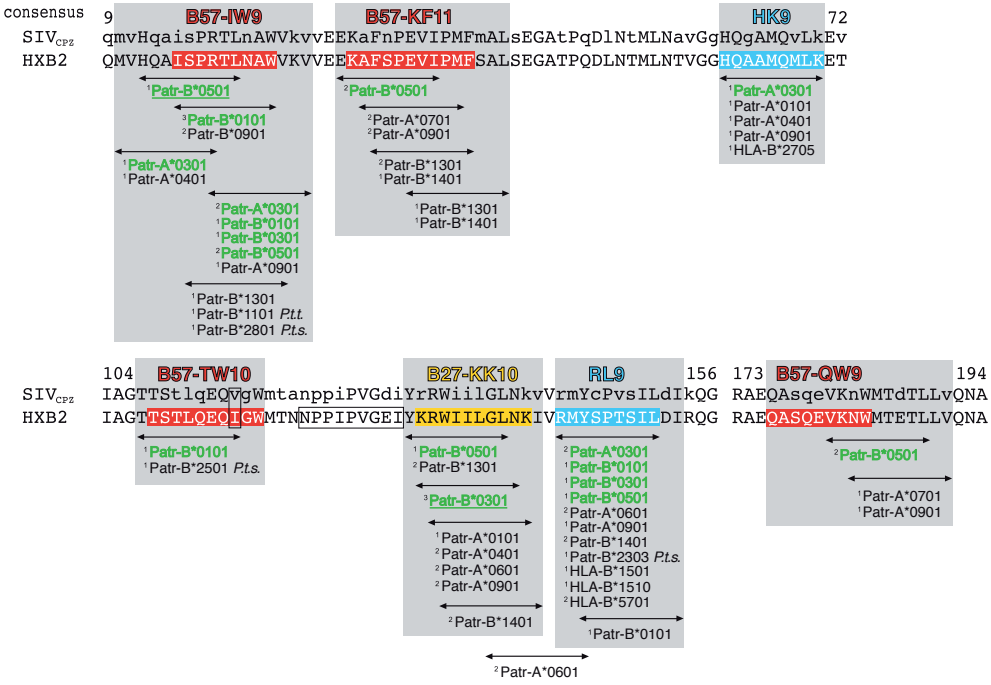


Fig. 2. Map of potential HIV-1 Gag CTL epitopes recognized by Patr class I molecules. The Gag sequence of HXB2 is taken as consensus, with the consensus of SIV_{cpz} above (www.hiv.lanl.gov). A lower-case letter in the SIV_{cpz} consensus indicates a variable position. HLA-B*27 and -B*57 CTL epitopes are marked in yellow and red, respectively. Potential 9-mer epitopes are indicated by an arrow, with the relevant Patr molecule(s) mentioned. Molecules marked in green indicate the predicted MHC/peptide combinations that are tested in peptide binding studies. Molecules specific for Central- (*P.t.troglodytes*, *P.t.t.*) or East-African chimpanzees (*P.t.schweinfurthii*, *P.t.s.*) are indicated. The predicted binding affinities using the NetMHCpan 2.0 algorithm are indicated by: [1]high, [2]intermediate, or [3]low. The arrow above the underscored Patr molecules indicate also previously reported CTL epitopes (30). The two potential epitopes, HK9 and RL9, identified with the NetMHCpan 2.0 algorithm are marked in blue. The control peptide, NPPIPVGEI, is boxed. The box in the B57-TW10 epitope indicates covered variation (Fig. S2).

of the SCs for the other three molecules. Because the peptides RTLNAWVKV and KRWIILGLN bound with a similar high affinity to Patr-B*0301 as its SC, this implicates a high binding affinity for these peptides as well. Moreover, for HK9, tested for Patr-A*0301, we observed a binding affinity similar to that of the SC, whereas RL9 has disparate binding affinities for the different molecules.

In addition, the selected peptides were tested in all other conceivable combinations with the four Patr molecules. The results, derived from pooled data from three individual experiments, are summarized (Fig. 3 and Fig. S2B). Based on a peptide binding motif,

Peptide	IC_{50} (μM)			
	A*0301	B*0101	B*0301	B*0501
SC	**1.5**	**1.0**	**0.2**	**1.3**
MVHQAISPR	**1.6**	3.6*	12.8*	0.3*
HQAISPRTL	3.9*	1.8	7.9*	**0.5***
ISPRTLNAW	0.7	**0.9**	2.7*	0.09*
RTLNAWVKV	**0.07***	**0.1***	**0.2**	**0.2***
EKAFSPEVI	18.1*	21.1*	>100*	**4.1***
TTSTLQEQV	9.1*	**5.0***	62.8*	2.9
TTSTLQEVI	6.3*	**9.6***	24.9*	5.2*
VKNWMTETL	0.1*	0.4	5.8*	**0.3***
YKRWIILGL	0.5*	1.3	0.3	**0.7**
KRWIILGLN	0.2*	0.6	**0.5**	0.03*
HQAAMQMLK	**2.3**	5.7*	>100*	1.1
RMYSPTSIL	**0.2***	**1.4**	**0.3**	**0.2***
NPPIPVGEI	**5.5***	**27.1***	**28.7***	**9.3***
EEALQAFTY	24.8*	22.9*	>100*	21.2*

Fig. 3. IC_{50} (μM) values of the different competitor peptides tested for Patr-A*0301, -B*0101, -B*0301, and -B*0501. The IC_{50} values in bold were determined from regression curves (Fig. S3) derived from pooled data from five individual experiments. For all other conceivable combinations, the IC_{50} values were derived from pooled data from three individual experiments. Previously described CTL epitopes are underscored (30). The color codes of the peptides correspond to the colors in Fig. 2. SC stands for standardized competitor (a peptide similar to the biotin-labeled indicator peptide). An asterisk indicates a significantly higher or lower binding affinity than the SC.

one would expect that each MHC class I molecule can bind a restricted spectrum of peptides. Indeed, we found that for most additional tested Patr/peptide combinations, the peptides bound with a lower binding affinity then that of the corresponding SC. However, for a few Patr/peptide combinations, a higher binding affinity then that of the corresponding SC was observed: for example, Patr-A*0301 in combination with KRWIILGLN (Fig. 3). Such additional specific binding was not predicted based on the peptide binding motifs. Apparently, however, some of the Patr class I molecules can bind a broader range of peptides that map to conserved areas of Gag. The molecular basis of this phenomenon is not yet understood. Finally, the two control peptides were found to bind with a low binding affinity to all four Patr molecules.

In conclusion, the data illustrate that Patr-B*0301 and -B*0501, each previously described to act as restriction elements for peptides in the KK10 and IW9 areas (30), respectively, also have the capacity to bind other peptides mapping to areas targeted by HLA-B*27/B*57. Overall, the four Patr molecules investigated seem to be able to bind peptides with a high binding affinity mapping to areas that overlap the IW9, KK10, and QW9 CTL epitopes. Hence, the majority of the peptides that were selected as potential epitopes indeed represent good binders for the respective molecules.

Evaluation of CTL data

The actual binding of a peptide to a MHC molecule does not warrant that immune recognition by effector cells will take place. On the basis of the present findings, we re-evaluated the cellular immune responses for three chimpanzees (Laurens, Zeef, and Thijs) belonging to an HIV-1-positive cohort (31). For Laurens, polyclonal CTL responses to four different Gag-pools were detected (Fig. 4). The highest percentage of specific lysis was detected for Gag-pool 1-3, and was at that time ascribed to epitope HQAISPRTL (part of the IW9 area) presented in the context of Patr-B*0501 (30). However, the present data illustrate that other molecules expressed by Laurens (Patr-A*0301 and - B*0101) can bind peptides overlapping with HQAISPRTL or the KF11 epitope, and could therefore contribute to the polyclonal CTL response to Gag-pool 1-3 (Fig. 4). Moreover, the molecules expressed by Laurens were found to bind the RL9 peptide and peptides overlapping the TW10, KK10, and QW9 epitopes, which may contribute to the polyclonal CTL responses observed for Gag-pools 11-14 and 15-18 (Fig. 4). For Zeef, only a response to Gag-pool 11-14 was observed (Fig. 4). Initially this response was ascribed to KRWIILGLN (part of the KK10 area) presented in the context of Patr-B*0301 (30). The results suggest that the combination of Patr-B*0101 and peptide TTSTLQEQI (part of the TW10 area) could contribute to the high specific lysis observed for peptide pool 11-14 as well. Finally, for Thijs, responses to three different Gag-pools were detected (Fig. 4), though at that time the restriction elements were not understood. Thijs does not possess Patr-B*0301 and -B*0501, responsible for the CTL responses in Zeef and Laurens, respectively. This implies that other molecules (for instance Patr-B*0101) are also able to recognize conserved areas (Fig. 4).

The absence of CTL responses to certain Gag-pools in Laurens, Zeef, and Thijs is most likely explained by the fact that these areas do not contain peptides that can bind to the molecules present in these particular animals. Alternatively, CTL-precursor frequencies may be too low to detect in the polyclonal assays used. In contrast, these areas can be recognized by various HLA molecules (www.hiv.lanl.gov). Thus, these results suggest that many of the CTL responses observed in chimpanzees may be directed to conserved areas of the Gag protein, which are also targeted by the HLA-B*27/B*57 molecules in humans.

Peptide binding results in the context of Patr class I haplotypes

A haplotype is defined as the unique combination of MHC alleles segregating together on a chromosome. MHC region genes are co-dominantly expressed, and like humans, chimpanzees possess one *A* and one *B* gene per haplotype. The peptide binding study results have been superimposed on the chimpanzee haplotypes. First, it was noted that most animals possess at least one molecule with the ability to bind a conserved area of Gag that is targeted by HLA-B*27 and/or -B*57 in humans (Fig. 5). Second, some animals encode multiple class I molecules, in cis- (e.g., Jacob and Yoko) or trans-configuration (e.g., Jolanda, Liesbeth, and Marco), which can bind similar and/or disparate conserved HIV-1 Gag peptides as HLA-B*27/B*57. These observations imply that the quality of the Patr molecule to bind conserved areas of Gag may play an important role in the protection against developement of AIDS.

Laurens

Gag Peptide pool	1-3	4-6	7-10	11-14	15-18	19-22
CTL response	+ (52%)			+ (8%)	+ (5%)	+ (8%)
Patr-A*0301	MVHQAISPR			RMYSPTSIL	RMYSPTSIL	
	RTLNAWVKV					
Patr-B*0101	ISPRTLNAW			TTSTLQEQI	RMYSPTSIL	
	RTLNAWVKV			RMYSPTSIL		
Patr-B*0501	HQAISPRTL			YKRWIILGL	VKNWMTETL	
	RTLNAWVKV			RMYSPTSIL	RMYSPTSIL	
	EKAFSPEVI					
Patr-A*0901	RTLNAWVKV			RWIILGLNK	RMYSPTSIL	
	AFSPEVIPM			RMYSPTSIL		

Zeef

Gag Peptide pool	1-3	4-6	7-10	11-14	15-18	19-22
CTL response				+ (55%)		
Patr-B*0101				TTSTLQEQI		
				RMYSPTSIL		
Patr-B*0301				KRWIILGLN		
				RMYSPTSIL		
Patr-A*0101				RWIILGLNK		
Patr-A*0601				RWIILGLNK		
				GLNKIVRMY		
				RMYSPTSIL		

Thijs

Gag Peptide pool	1-4	5-8	9-12	13-16	17-19	20-22
CTL response	+ (18%)	+ (12%)		+ (35%)		
Patr-B*0101	ISPRTLNAW			RMYSPTSIL		
	RTLNAWVKV					
Patr-A*0401[a]	MVHQAISPR	HQAAMQMLK		RWIILGLNK		
Patr-B*1401	FSPEVIPMF			WIILGLNKI		
	EVIPMFSAL			RMYSPTSIL		

Fig. 4. Analysis of cellular immune responses in three chimpanzees belonging to an experimentally infected HIV-1 cohort. The peptides present in the Gag peptide-pools are summarized in Fig. S4. + Indicates that a CTL response to the Gag-pool was observed, and the percentage-specific lysis is given in brackets (31). The Patr-A and -B molecules present on the haplotypes of chimpanzees Laurens, Zeef, and Thijs are given. The relevant MHC/peptide combinations that probably contribute to the previously observed Gag-pool-specific CTL responses are given, and the color codes of the peptides correspond to the colors in Fig. 2. Previously described CTL epitopes are underscored (30). A grey background indicates data obtained using only the NetMHCpan 2.0 algorithm. [a]Animal is homozygous for this molecule.

However, the fact that particular animals encode multiple class I molecules able to bind conserved Gag peptides suggest that in chimpanzees perhaps also a quantitative feature is operative. In conclusion, the majority of the founder animals possess at least one Patr molecule that, like HLA-B*27/B*57, is able to bind peptides from similar conserved areas of Gag, and as such could respond to an HIV-1/SIV$_{cpz}$ infection in a similar way as described for HLA-B*27/B*57-positive human long-term non-progressors.

Chimpanzee	A*		B*	
Carolina	0401	NT	0101	NT
Debbie	0901	0101	0101	1401
Diana	0501	0302	0101	1301
Frits	0601	0401	0301	1401
Gerrit	0501	0301	1301	2901
Gina	0901	NT	0101	1401
Igor	0701	NT	2402	NT
Indira	0101	0701	1401	0501
Isaac	0101	1401	0401	0101
Jacob	0301	0501	0501	1301
Jolanda	0901	0301	0101	2901
Karin	0701	1101	0501	1301
Lady	NT	0301	NT	2402
Liesbeth	0701	0301	0101	2402
Louise	0602	0101	0901	0101
Marco	0901	NT	0101	0501
Marga	NT	0101	0501	16011
Mario	0701	NT	0301	1401
Nina	0901	0601	0101	0501
Pearl	1101	0601	1301	0101
Regina	0401	0201	0101	1401
Renee	0901	1401	0101	0101
Renza	NT	0701	1301	NT
Sherry	0701	1401	0501	0101
Sonja	0401	0901	0101	0101
Susie	0301	NT	2402	NT
Tasja	0301	0501	1301	2402
Tineke	0401	NT	0101	NT
Toetie	0901	0301	0101	2402
Wodka	NT	0101	0501	1401
Yoko	0301	1401	0501	16011
Yvonne	0301	NT	2402	NT

Fig. 5. MHC class I haplotypes of the founder chimpanzees. Molecules that bind conserved Gag peptides overlapping the areas targeted only by HLA-B*57 are presented in red, whereas orange indicates molecules binding peptides overlapping the areas targeted by HLA-B*27 and -B*57. Boxed molecules are predicted to bind HLA-B*27/B*57-like overlapping conserved Gag peptides based on the NetMHCpan 2.0 algorithm. NT means not typed, because data are no longer accessible.

Discussion

We have investigated whether the selective sweep targeting the Patr class I repertoire was possibly caused by HIV-1/SIV$_{cpz}$ (or a closely related retrovirus) and has resulted in the preferential selection of molecules that confer resistance to AIDS. The peptide binding motifs of four frequent Patr class I molecules were defined, and we found that these molecules, like the AIDS-protective HLA-B*27 and -B*57 molecules, bind peptides derived from conserved areas of the HIV-1/SIV$_{cpz}$ Gag-protein. 94% of the wild-caught West-African chimpanzees studied possess at least one of these four Patr class I molecules. Moreover, many chimpanzees express several molecules that can bind multiple peptides originating from various conserved Gag regions, which suggests that chimpanzees may have developed a "double-lock" strategy. In humans the importance of heterozygous advantage with regard to AIDS resistance is also applicable (32), and one would expect that heterozygous individuals that contain both, HLA-B*27 and -B*57, progress more slowly to AIDS or behave as elite controllers (33). Accordingly, three chimpanzees of an experimentally infected HIV-1 cohort were shown to display broadly reactive CTL responses to conserved areas of HIV-1. Taken together, the results suggest that the chimpanzee MHC class I repertoire was skewed to favor the survival of animals whose Patr molecules recognized conserved areas of Gag, a property shared with HLA-B*27/B*57. This may be due to the sharing of similar peptide binding motifs between human and chimpanzees but also to the ability of some Patr molecules to target overlapping areas of the Gag protein, which are conserved across different strains. As such, the present report significantly extends the information that was gathered in an earlier CTL study (30). Apart from Patr-B, also a Patr-A molecule was shown to bind peptides from conserved Gag areas similar to those targeted by HLA-B*27/B*57, and as such may contribute to controlling HIV-1/SIV$_{cpz}$ viral load. In comparison to HLA-A, Patr-A molecules have more promiscuous peptide binding motifs, allowing the presentation of a greater spectrum of peptides (Fig. 1). Nonetheless, one should bear in mind that we only determined the peptide binding motifs for four Patr molecules. Whether the picture that emerges may become more prominent could be addressed when peptide-binding and cellular data become available for other molecules. To verify this assumption, the Gag protein was scanned with the NetMHCpan 2.0 algorithm (28) for two additional molecules: Patr-A*0401 and -A*0901. High- or intermediate-affinity peptides overlapping the IW9 and KK10 epitopes were identified (Fig. 2), and indeed CTL activity to Gag-pools covering these peptides was observed (Fig. 4). The incorporation of the data for Patr-A*0401 and -A*0901 into the haplotype list further strengthen both the qualitative and quantitative feature of the immuneresponse to respond to an HIV-1/SIV$_{cpz}$ infection (Fig. 5). However, it has to be taken into consideration that it is likely that other Patr molecules may not be able to target conserved areas of Gag similar to those targeted by HLA-B*27/B*57.

Immunodominant CTL responses to Gag found in HLA-B*27/B*57-positive human long-term non-progressors and HIV-1 infected chimpanzees may play a crucial role in control

of viral replication (19, 27, 30). The Gag protein may represent one of the Achilles heels of HIV-1/SIV_{cpz}. However, from a mechanistic point of view it is not clearly understood how this works (27), since not all HLA-B*27 or -B*57-positive individuals can control an HIV-1 infection. Although heterozygous advantage may have an additional impact, other host and viral factors may play a role as well. In this respect, it is worth to mention that we do not want to claim that the MHC is the only genetic region that explains the relative resistance to AIDS in chimpanzees, since other genes, such as for the chemokine (C-C motif) receptor 5 and CD4, show signs of selection and may have been affected by a sweep as well (34, 35). Furthermore, it was shown recently that rhesus macaques that are elite controllers in AIDS vaccine studies possess T cell responses to different viral proteins, primarily Vif and Nef, restricted by the Mamu-B*08 or -B*17 molecule (36, 37). These observations illustrate that at least in rhesus macaques T cell responses to different viral proteins could play a prominent role in elite control. Therefore, the present study, in which we focussed on different conserved Gag epitopes, may only represent a part of our understanding how chimpanzees may control viral replication. Follow-up studies are needed to sort out the role of other viral proteins in controlling development of AIDS in chimpanzees.

A recent report indicated that naturally infected East-African animals contract AIDS (4). The fact that chimpanzees can develop AIDS was shown earlier for one animal that was found to contain a recombinant virus that emerged after a co-infection with three different HIV-1 isolates (7). After transfer of this virus isolate to other chimpanzees, it caused $CD4^+$ T-cell depletion and increased plasma viral loads as well, indicating that specific HIV-1 isolates are able to provoke AIDS or AIDS-like symptoms. SIV_{cpz} strains isolated from naturally infected chimpanzees appear to represent a mosaic of Old World monkey virus segments (38). Chimpanzees and their ancestors were most likely infected by predating on different monkey species known to have been infected with several types of SIV strains. It is plausible that new recombinant SIV_{cpz} strains are generated occasionally (39), and that some of these strains have pathogenic characteristics. Non-human primate species have been challenged by SIV infections over long periods of time, and must have developed ways to control the development of AIDS (15, 40), like the AIDS-resistant West-African chimpanzees studied here. There is evidence that the selective sweep or subsequent selection processes must have been more prominent in West-African chimpanzees than in other chimpanzee populations (34, 35, 41). This may be due to population separation, and may be dependent on the monkey species being predated upon and on their respective SIV infections (42). As a consequence, repertoires could have been edited in slightly different manners. Nevertheless, the broad picture that is emerging suggest that most chimpanzees possess MHC class I molecules that bind conserved epitopes in Gag, similar to those targeted by AIDS-resistant HLA-B*27/B*57-positive humans. The present functional characteristics of the skewed MHC class I repertoire in chimpanzees suggest that the ancient selective sweep was caused by a lentiviral pandemic. As a consequence most chimpanzees seem to be able to cope with retroviral infections such as HIV-1/SIV_{cpz}, like AIDS-resistant HLA-B*27/B*57-positive humans.

Materials and Methods

Patr-A and -B transfected cell lines

Constructs containing the relevant *Patr-A* or *-B* sequences were individually transfected into K562 cells (which lack HLA-expression) by electroporation. The details are described in *SI Materials and Methods*.

Determination of peptide binding motifs

K562 *Patr-A* or *-B*-positive transfectants were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with G418 and 5% fetal calf serum (FCS) up to $1x10^{10}$ cells. Subsequently, the cells were lysed. Sepharose beads covalently linked with monoclonal antibody W6/32 were used to pre-clear the lysate. The beads were then washed, and in a final step the MHC-peptide complexes were eluted with 10% acetic acid in water. High molecular mass material (MHC molecules) was removed through Centriprep filtration. The peptide pool was prefractionated on a C18 RP-HPLC system (Dr. Maisch GmbH, Ammerbuch, Germany), and was subsequently analyzed by nanoHPLC-tandem mass spectrometry with a LTQ-FT. The details are described in *SI Materials and Methods*.

Selection and synthesis of indicator and competitor peptides

The details are described in *SI Materials and Methods*.

Cell-based peptide binding competition assays (CPBCA) and CTL assays

A plate-based CPBCA was developed to measure the binding affinity of the different peptides using time-resolved fluorescence as read-out (see also Fig. S5). The full protocol is described in *SI Materials and Methods*. Previously described CTL assays were re-evaluated (30, 31).

Calculation of percentage inhibition, IC_{50} values, and statistical analysis

The details are described in *SI Materials and Methods*.

Acknowledgements

The authors wish to thank D. Devine for editing the manuscript, H. van Westbroek for preparing the figures, Drs J. Bajramovic, D. Davis, and G. Koopman for their suggestions and critical reading of the manuscript, Prof. Dr. P. van den Elsen of the Leiden University Medical Center, The Netherlands for kindly providing the pcDNA/neo vector and MC1061/P3 E. coli, Dr. B. Faber for purification and concentration of the monoclonal antibody W6/32, and Dr. D. Stepniak for assistance with the peptide isolation. This study was supported by the Biomedical Primate Research Centre.

References

1. Fujiyama A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
2. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287:607-614.
3. Santiago ML, et al. (2002) SIVcpz in wild chimpanzees. Science 295:465.
4. Keele BF, et al. (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460:515-519.
5. Heeney JL, et al. (2006) Transmission of simian immunodeficiency virus SIVcpz and the evolution of infection in the presence and absence of concurrent human immunodeficiency virus type 1 infection in chimpanzees. J Virol 80:7208-7218.
6. Heeney JL, Dalgleish AG, Weiss RA (2006) Origins of HIV and the evolution of resistance to AIDS. Science 313:462-466.
7. Novembre FJ, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
8. Kirchhoff F, et al. (1995) Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. N Engl J Med 332:228-232.
9. Samson M, et al. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722-725.
10. Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67-74.
11. Kiepiela P, et al. (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 432:769-775.
12. Martin MP, et al. (2007) Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nat Genet 39:733-740.
13. Dalmasso C, et al. (2008) Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS ONE 3:e3907.
14. Fellay J, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317:944-947.
15. de Groot NG, et al. (2005) Reduced MIC gene repertoire variation in West African chimpanzees as compared to humans. Mol Biol Evol 22:1375-1385.
16. de Groot NG, et al. (2008) Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. Mol Ecol 17:2074-2088.
17. de Groot NG, et al. (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci U S A 99:11748-11753.
18. Yohn CT, et al. (2005) Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. PLoS Biol 3:e110.
19. Borghans JA, Molgaard A, de Boer RJ, Kesmir C (2007) HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. PLoS ONE 2:e920.
20. Martinez-Picado J, et al. (2006) Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J Virol 80:3617-3623.
21. Rolland M, et al. (2008) Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. PLoS ONE 3:e1424.
22. Falk K, et al. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290-296.

23. Sidney J, et al. (2006) Detailed characterization of the peptide binding specificity of five common *Patr* class I MHC molecules. Immunogenetics 58:559-570.
24. Sidney J, et al. (2007) Characterization of the peptide-binding specificity of the chimpanzee class I alleles *A*0301* and *A*0401* using a combinatorial peptide library. Immunogenetics 59:745-751.
25. Marsh SGE, Parham P, Barber LD (2000) The HLA Factsbook (Academic Press, San Diego).
26. Rammensee H, et al. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213-219.
27. Goulder PJ, Watkins DI (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol 8:619-630.
28. Hoof I, et al. (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61:1-13.
29. Perez CL, et al. (2008) Broadly immunogenic HLA class I supertype-restricted elite CTL epitopes recognized in a diverse population infected with different HIV-1 subtypes. J Immunol 180:5092-5100.
30. Balla-Jhagjhoorsingh SS, et al. (1999) Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees. J Immunol 162:2308-2314.
31. Balla-Jhagjhoorsingh SS, et al. (2001) Protection from secondary human immunodeficiency virus type 1 infection in chimpanzees suggests the importance of antigenic boosting and a possible role for cytotoxic T cells. J Infect Dis 184:136-143.
32. Carrington M, et al. (1999) HLA and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. Science 283:1748-1752.
33. Bailey JR, et al. (2008) Transmission of human immunodeficiency virus type 1 from a patient who developed AIDS to an elite suppressor. J Virol 82:7395-7410.
34. Hvilsom C, et al. (2008) Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene. Genomics 92:322-328.
35. MacFie TS, et al. (2009) Patterns of diversity in HIV-related loci among subspecies of chimpanzee: concordance at CCR5 and differences at CXCR4 and CX3CR1. Mol Biol Evol 26:719-727.
36. Mothe BR, et al. (2002) Characterization of the peptide-binding specificity of Mamu-B*17 and identification of Mamu-B*17-restricted epitopes derived from simian immunodeficiency virus proteins. J Immunol 169:210-219.
37. Valentine LE, et al. (2009) Infection with "escaped" virus variants impairs control of simian immunodeficiency virus SIVmac239 replication in Mamu-B*08-positive macaques. J Virol 83:11514-11527.
38. Bailes E, et al. (2003) Hybrid origin of SIV in chimpanzees. Science 300:1713.
39. Leitner T, et al. (2007) Sequence diversity among chimpanzee simian immunodeficiency viruses (SIVcpz) suggests that SIVcpzPts was derived from SIVcpzPtt through additional recombination events. AIDS Res Hum Retroviruses 23:1114-1118.
40. Sodora DL, et al. (2009) Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. Nat Med 15:861-865.
41. Wooding S, et al. (2005) Contrasting effects of natural selection on human and chimpanzee CC chemokine receptor 5. Am J Hum Genet 76:291-301.
42. Liegeois F, et al. (2009) Full-length genome characterization of a novel simian immunodeficiency virus lineage (SIVolc) from olive Colobus (*Procolobus verus)* and new SIVwrcPbb strains from Western Red Colobus (*Piliocolobus badius badius*) from the Tai Forest in Ivory Coast. J Virol 83:428-439.

43. McKinney DM, et al. (2000) Identification of five different *Patr* class I molecules that bind HLA supertype peptides and definition of their peptide binding motifs. J Immunol 165: 4414-4422.

Supplementary material

The following supplementary material is available for this article:

SI Material and Methods.
SI References.

Fig. S1. Peptide binding motifs determined from the eluted natural peptides. (A) Patr-A*0301, (B) -B*0101, (C) -B*0301, and (D) -B*0501. WebLogo version 3.0 was used for visualization of the eluted 9-mer peptides (9). In a sequence logo, the height of a column of letters is equal to the information content at that position. The height of a letter in a column is proportional to the frequency of the corresponding amino acid at that position. The color code of the amino acids is according to the chemical properties: green for polar, blue for basic, red for acidic, and black for hydrophobic amino acids. As an example, a list of eluted 8-, 9-, and 10-mer natural peptides used to elucidate the peptide binding motif for Patr-A*0301 is provided (Table S2).

Fig. S2. Overview of the indicator and competitor peptides tested for JY, DBB, HLA-A2 (SAL), and the different Patr molecules. The determined IC_{50} values and the corresponding 95% confidence intervals (CI) are provided. (A) The IC_{50} values for JY/DBB/HLA-A2(SAL) and the Patr molecules are the result of three and five individual experiments, respectively. For the Patr molecules, the data of both tested indicator peptides are shown. (B) The IC_{50} values presented are the result of three individual experiments. **C** indicates the position where the indicator peptides are labeled. SAL means single molecule-expressing cell-line. [1]ATALEYVYK is a biogenesis of lysosome-related organelles complex-1 subunit 1. [2]LSDMHLRSI is an isoform 1 of fragile X mental retardation syndrome-related protein 1. [3]GRIDIKQLI is a Sterol O-acyltransferase 1. [4]GQYEQVKQL is a CCR4-NOT transcription complex, subunit 1 isoform b. [5]EEALQAFTY is an alanine substituted peptide of the ATP- binding cassette sub-family D member 3 derived peptide EEYLQAFTY (10). The color codes correspond to those in Fig. 2.

Fig. S3. Individual dose-inhibition curves for the competitor peptides predicted to bind tested for (A) Patr-A*0301, (B) -B*0101, (C) -B*0301, and (D) -B*0501. The indicated IC_{50} values (μM) are derived from the regression curves of five individual experiments. As indicator peptides ATALE**C**VYK biotin-labeled at the p6 position (indicated with a C) and LSDMHL**C**SI, GRIDIK**C**LI, and GQYEQV**C**QL biotin-labeled at the p7 position were used, respectively. A triangle marks the standardized competitor (SC; peptide similar to the biotin-labeled indicator peptide). A square indicates the Gag-derived control peptide. An asterisk indicates a binding affinity significantly higher or lower than the SC.

Fig. S4. Individual peptides present in the Gag-pools of the cellular immune response assays. Twenty-two 20-mer peptides overlapping by ten amino acids spanning the HIV-1$_{SF2}$ (ARP-788; MRC) Gag amino acid residues 135-364 (8).

Fig. S5. Dose-inhibition curves for (A) JY, (B) DBB, and (C) HLA-A*02 (SAL). As indicator peptide HBV cAG 18-27 biotin-labeled at the p6 position was used. The blue lines represent competition with unlabeled HBV cAG 18-27 (standardized competitor peptide). The red lines represent competition with HIV-1 p24 15-23, a pepitde predicted to have no/low binding affinity. The indicated IC_{50} values (μM) are derived from the regression curves of three individual experiments. An asterisk indicates a binding affinity significantly lower than the standardized competitor peptide.

Table S1. Numbers of 8-, 9-, and 10-mer natural eluted peptides with a mascot value of 35 or higher for the different Patr-A and -B molecules.

Table S2. List of 8- (yellow), 9- (blue), and 10-mer (salmon) eluted natural peptides with a mascot value of 35 or higher for Patr-A*0301.

This material will become available as part of the online article.

Chapter 8

General Discussion:

The HIV-1 pandemic:
Do chimpanzees mirror mankind's future?

The HIV-1 pandemic

In 1983, the human immunodeficiency virus type 1 (HIV-1) was identified to have initiated the death of particular human individuals (1). The disease was given the name "acquired immunodeficiency syndrome" (AIDS) (2, 3). At present over thirty million people worldwide are infected with HIV-1, and more then twenty five million have died from AIDS since the beginning of the pandemic (UNAIDS, 2009). Thanks to the currently available antiretroviral (AR) drugs, the number of people dying yearly from AIDS has reached a plateau somewhere around 2004, and the number has even begun to decline (UNAIDS, 2009). But, due to the high cost of AR therapy, only a limited number of HIV-1-infected individuals have access to the drugs. Especially in developing countries, where most of the infected individuals live, AIDS is still a devastating disease. Therefore, the need for low-cost drugs, or even better, an efficient vaccine able to protect against infection, is obvious but is still lacking at present. Taking into consideration the failure of the Merck HIV-1 vaccine in 2007 (4), and the controversy regarding expectations concerning the RV144 HIV-1 vaccine trial started in 2003 (5), one may question what can be learned from non-human primates that are naturally infected with simian immunodeficiency virus (SIV).

Several African non-human primate species are naturally infected with various SIV strains, but progression towards AIDS is rarely observed. This is thought to have been due to an evolutionary adaptation between the host's immune system and the SIVs, which resulted in an equilibrium (6). Chimpanzees (*Pan troglodytes*, *Patr*) can also be naturally infected with SIV (7). This strain, SIV_{cpz}, is considered to be the initiator of the human HIV-1 pandemic, and genetic analyses revealed that the pandemic originated from at least three independent cross-species transmissions of SIV_{cpz} (8).

Chimpanzees are the only non-human primate species that is susceptible to infection with HIV-1, and they were therefore used as an experimental model to evaluate vaccines that could protect the human population against this lethal infection. Over 150 chimpanzees worldwide have been experimentally infected with different HIV-1 strains, but since they appeared to be relatively resistant to developing AIDS, studying AIDS pathology in chimpanzees was not possible (9). However, the infection of rhesus macaques (*Macaca mulatta*) with SIV_{mac} does result in AIDS-like disease (10-12), and this model is currently widely used to study immunopathogenesis as well as to modulate the immunological responses induced by HIV-1 vaccines or vaccine components.

AIDS resistance: A role for the MHC?

Several studies in HIV-1-infected human cohorts revealed that in certain individuals the infection does not progress to AIDS. Some of these so-called long-term non-progressors (LTNP)/elite controllers have been infected for over 17 years, and their situation appears to

resemble that of HIV-1-infected chimpanzees not progressing towards AIDS. Detailed analysis showed that the LTNP status was strongly associated with the presence of particular major histocompatibility complex (MHC) class I molecules, such as HLA-B*2705 and -B*5701 (13).

The classical MHC class I loci in humans and chimpanzees are designated *HLA-* and *Patr-A*, *-B*, and *-C*, respectively. In comparison to humans, chimpanzees have a reduced *Mhc* class I *A* repertoire. They only appear to possess orthologs to the *HLA-A1/A3/A11/A30* family, whereas orthologs to the other five *HLA-A* families are absent (14). As a follow up, we studied the MHC class I repertoire of a pedigreed West-African chimpanzee colony (**Chapter 2**). In this case as well, only orthologs to the *HLA-A1/A3/A11/A30* family were observed. The results demonstrated that, within the population analysed, a relative high degree of allelic variability was observed for each of the limited number of *Mhc* class I lineages that was detected. Nevertheless, though the results suggested that there has been a repertoire reduction at the MHC class I region, one should exercise caution. First, MHC class I and II gene products may experience different sorts of natural selection, and the possibility of convergent evolution cannot be excluded. Second, an imbalance in sample size between the number of humans and chimpanzees analyzed can hamper an accurate interpretation of the data. For instance, one can argue that alleles and/or lineages were not observed, due to a small sample size. Therefore, the influence of selection operating on the classical MHC class I loci was studied by comparing the intron variation in human and chimpanzee alleles in two well-defined populations (**Chapter 3**). This analysis revealed that chimpanzees indeed experienced a selective sweep targeting the MHC class I repertoire, which is most prominent for the B locus ($\chi^2 = 2.64$, $P = 0.01$), and which was claimed to predate the (sub)speciation of chimpanzees. In comparison to humans, chimpanzees are relatively resistant to the development of AIDS after natural SIV_{cpz} or experimental HIV-1 infection. In humans, AIDS resistance is strongly associated with the presence of particular MHC class I molecules. From an evolutionary perspective, chimpanzees and humans are each other's closest living relatives, which implies that a similar MHC repertoire must have been present in their common ancestor. Chimpanzees, however, possess a reduced MHC class I repertoire as the result of a selective sweep. The aforementioned observations resulted in the hypothesis that the selective sweep was caused by HIV-1/SIV_{cpz} or a closely related retrovirus. As a consequence, contemporary chimpanzee populations represent the offspring of AIDS-resistant animals, the survivors of an HIV-1-like pandemic that took place in the distant past (**Chapter 3**).

Genetic diversity beyond the MHC class I loci and its role in resistance/susceptibility to AIDS

Different gene products of the human immune system are involved in HIV-1 infection, and several of them may play a role in the resistance/susceptibility to AIDS. HIV-1 belongs to the

group of lentiviruses and possesses specific properties that have made the development of broadly efficacious and efficient anti-retroviral drugs and vaccines difficult. On the one hand, the virus can cause a persistent infection and hide in an inactive form in cellular reservoirs. On the other hand, it can replicate very quickly (10^9 to 10^{10} particles/day), and the error-prone reverse transcriptase has a high mutation rate ($3x10^{-5}$ per nucleotide base per cycle of replication), which may ultimately result in the generation of many variants (virus swarm) in a single infected individual.

HIV-1 enters the human body via the lower and upper mucosal sites, blood-blood contact, or breast milk. After entry, the virus either directly infects $CD4^+$ T cells or is taken up by dendritic cells (DC) or macrophages. These latter two cell types transport HIV-1 through different layers of epithelial cells to the lymphoid tissues, where they transfer the virus to $CD4^+$ T cells. DCs internalize HIV-1 into endosomes, and the responsible cell-surface molecule is the C-type lectin receptor DC-SIGN. Macrophages phagocytose HIV-1, and either function as a reservoir for long-term persistence of the virus or are responsible for the transport and dissemination of the virus. The "primary" non-syncytium-inducing (NSI) HIV-1 viruses mainly infect $CD4^+$ memory T cells that express CC-chemokine receptor-5 (CCR5). Later in the course of the infection, the NSI strains can switch to a syncytium-inducing (SI) phenotype, which can infect T-cells by using the C-X-C-chemokine receptor-4 (CXCR4) as co-receptor. This switch is associated with a loss of sensitivity to chemokines (for instance RANTES, and MIP-1α, and -β), a rapid decrease in $CD4^+$ T cells, and progression towards AIDS (15). In addition, an infection of resting T cells ensues as well, resulting in a dormant stage of the virus, which is undetectable by the host's immune system.

Immature DCs are one of the first cells that interact with HIV-1 at the site of infection, and the surface molecule DC-SIGN is thought to play an essential role in disseminating the virus (16). Polymorphisms in the repeat region of human DC-SIGN are associated with resistance to HIV-1 infection (17, 18). A homolog of human DC-SIGN is found in chimpanzees, and appears to have a similar function (19). Polymorphisms in the repeat region of chimpanzee DC-SIGN are described (19-21), but unlike in humans, the number of repeats could not be correlated to susceptibility to HIV infection (21).

Homologs of the human CD4 receptor, and of the CCR5 and CXCR4 co-receptors, are present in chimpanzees. Genetic analysis of the chimpanzee CD4 receptor revealed differences relative to its human counterpart. Chimpanzees are divided into four subspecies based on mitochondrial DNA (mtDNA) and geographic origin (22, 23). Individuals of the West-African subspecies (*P.t.verus*) have a conserved CD4 receptor, whereas the other subspecies harbor highly variable CD4 receptors (24). Further investigation is needed to determine whether the species-specific variations influence susceptibility to HIV-1/SIV_{cpz} infection (24).

In humans, several genetic CCR5 modifications are associated with resistance/susceptibility to HIV-1 infection (25). The CCR5 variant that possesses a 32-base pair deletion (CCR5-Δ32) confers nearly complete resistance to HIV-1 infection in homozygous individuals, and is present in approximately 1% of the Caucasian population (26). The chimpanzee CCR5 and CXCR4 genes are highly homologous to the human ones, and no 32-base-pair deletion in

the chimpanzee CCR5 was observed that could explain the lack of progression towards AIDS (27-29). Both receptors, CCR5 and CXCR4, can be used by HIV-1/SIV_{cpz} to gain entry into chimpanzee cells. A recent study has demonstrated patterns of diversity for the CCR5, CXCR4, and CX3CR1[1] loci in three chimpanzee subspecies (*P.t.troglodytes*, *vellerosus*, *verus*) by using microsatellite markers (31). The CCR5 locus showed a low diversity for all three subspecies, which suggests a selective sweep at this locus that took place before subspeciation. For CXCR4 and CX3CR1, other evolutionary forces seem to be responsible for the observed subspecies diversity. For *P.t.troglodytes*, natural infections with SIV_{cpz} are observed, and a selective sweep at the CCR5 locus may be related to co-evolution with SIV_{cpz}. The fact that *P.t.vellerosus* and *verus* also show evidence for selection at CCR5, but natural infections with SIV_{cpz} are not observed thus far (32), could suggest that these subspecies were infected with a SIV_{cpz}-like virus in the past, and that infections are rare or absent at present. Alternatively, the ancestral chimpanzee lineage may have been infected with SIV_{cpz}, resulting in a prominent selection event at that stage (31).

Variations in the human 5'CCR5 region are associated with different transcription levels that influence HIV-1 entry and may affect disease progression. Chimpanzees underwent directional selection for the 5'CCR5 region, and the present functional variants may confer resistance to HIV-1/SIV_{cpz}-induced disease in chimpanzees (33). Hence, the human 5'CCR5 haplotype that shows the lowest promoter activity resulting in AIDS resistance is the most common haplotype in chimpanzees (33, 34).

Activated $CD4^+$ memory T cells are selectively and rapidly depleted after HIV-1/SIV infection (35-37). This cell type is abundantly present in the intestine, but other mucosal surfaces (lung, vagina) that are frequently exposed to environmental antigens also contain a high percentage of activated lymphocytes (38). Studies in humans and macaques have illustrated that in particular the intestine is the earliest site of $CD4^+$ T cell depletion during the first few weeks after exposure to HIV-1/SIV. Highly expressed on the gastrointestinal epithelium is the major histocompatibility complex class I chain-related gene (*MIC*) molecule (39). In humans, seven genes (*MICA* to *MICG*) have been distinguished (40), though only the *MICA* and *B* genes produce functional transcripts and display high levels of polymorphism (41). Chimpanzees have one functional *MIC* gene (42), which has an intermediate character between the human *MICA* and *B* genes (43). A large deletion of 95 kb has resulted in a *Patr-MICA/B* fusion product (44), but a thorough population study is lacking, and only one allele has been documented (42). *MIC* genes are located in the MHC region, with the *Patr-MICA/B* gene situated near the *Patr-B* locus (45), the one that showed the most severe repertoire reduction in chimpanzees (**Chapter 3**). To examine whether the chimpanzee MHC repertoire reduction affected other genes located in the MHC region, we investigated *MIC* gene polymorphism in the West-African chimpanzee population (**Chapter 4**). All haplotypes sampled possess the *Patr-MICA/B* fusion gene, and only one lineage showing moderate allelic variation was observed.

[1]Chemokine (C-X3-C motif) receptor 1, polymorphisms in the receptor can influence the susceptibility to HIV-1 infection (30).

In conclusion, the reduced *Patr-MIC* gene repertoire is compatible with the selective sweep observed for the *Patr-B* locus (**Chapter 4**).
In humans, MIC is the ligand for the NKG2D receptor, which is expressed on NK cells, γδ T cells, and $CD8^+$ αβ T cells (46-48). Cellular stress, triggered for instance by a viral infection or a malignant transformation, upregulates the expression of MIC, which can ultimately lead to an immune response (46, 47). The Patr-MICA/B fusion molecule recognizes human Vδ1 γδ T cells specific for MICA and B, suggesting a conserved recognition site (42). In addition, the NKG2D receptor is also highly similar between humans and chimpanzees (49). Whether MIC plays a role in NK effector responses against HIV-1/SIV_{cpz} is yet to be proven, as is the functional role of the *Patr-MICA/B* gene in viral infections (**Chapter 4**).

Comparative genomics between humans and chimpanzees: A different approach to specify the region targeted by the selective sweep

Approximately 5-6 million years ago (MYA), humans and chimpanzees shared a common ancestor, a fact that is still reflected in a 98.7% similarity at the nonrepetitive DNA level (50). Analysis of mtDNA, which is only inherited via the maternal line, shows that great apes display more variation than humans (22, 51). The different chimpanzee subspecies shared a common ancestor approximately 1.5 MYA, whereas the modern human lineage have existed for 150,000-200,000 years (52). This indicates that the chimpanzee as a species is subjected to a different time scale with regard to generating genetic variation: this is not only reflected in more mtDNA variation but also in more variation in particular nuclear genes (53, 54). The West-African chimpanzee population studied represented an outbred population, showing abundant mtDNA variation (**Chapters 4 and 5**).
For evolution and population studies, microsatellites, which are short tandem repeats of 2 to 6 base pairs, are valuable markers (55, 56), and can be used to evaluate whether a genetic region is affected by a repertoire reduction, such as a selective sweep (57). We compared the genetic variation of microsatellite markers located in the MHC region with unlinked autosomal markers mapping outside the MHC region (non-MHC) to define more specifically the segment and also the extent of the region that was targeted by the selective sweep (**Chapter 5**). Based on the assumption that chimpanzees are older as a species than humans, and thus had more time to generate genetic variation, one would expect to find a more diverse or at least an equivalent complex microsatellite profile in chimpanzees. In keeping with this expectation, we found similar amounts of variation for the non-MHC markers. The expected heterozygosity index (H_E) was on average 0.63 ± 0.15 in humans and 0.61 ± 0.20 in chimpanzees. In contrast, most studied MHC markers showed a lower allelic variation in chimpanzees than in humans, with an average H_E of 0.52 ± 0.27 versus 0.78 ± 0.11. This indicated that, compared to humans, chimpanzees have a reduced genetic repertoire in the MHC class I and II regions. However, most chimpanzees studied do possess unique MHC haplotypes, although the animals were sampled from one single population originating from Sierra Leone.

Most likely, the sweep selected some of the haplotypes, and particular segments were reshuffled by naturally occurring recombination-like processes contributing to the recovery of MHC haplotype diversity. Nevertheless, these processes did not have sufficient time to erase the signal of the selective sweep (**Chapter 5**).

In addition, multilocus demographic analyses for the MHC and non-MHC markers within the chimpanzee samples were performed to test whether the genetic MHC reduction is due to selection and/or demographic history. One has to bear in mind that it is difficult to date an ancient selective sweep using microsatellite data, due to their high mutation rate. The analyses revealed that independent of the demographic model assumed, linear or exponential, the chimpanzee population decreased from at least 10,000 to around 100 effective numbers of individuals. This decrease was dated between >3000 and <10,000 years for the exponential model, and between 10,000 and 100,000 years for the linear model. The linear model of population decrease assumed a more ancient decrease for the MHC than for the non-MHC data, which may reflect a signature of the ancient selective sweep. In addition, the single locus estimation revealed that the variance for the non-MHC markers reflects the genealogical stochasticity, whereas the MHC markers could be divided into two groups: five markers in the MHC class I region with an ancestral population size, log (N_1), of five, and four markers in the class II region with a log (N_1) of approximately four. The current population size, log (N_0), of approximately two is the same for all MHC markers. This underscores the likelihood that genetic variation in the MHC class I region has been reduced mostly by selection. The multilocus demographic analyses indicated that chimpanzees experienced a selective sweep that mainly targeted the chromosomal segment carrying the MHC class I region. Other polymorphic loci, like the MHC class II loci and MIC, mapping in the close vicinity of the MHC class I region genes, are probably affected due to genetic linkage (**Chapter 5**).

HIV-specific CD4$^+$ T helper cells in host defense: The importance of MHC class II

CD4$^+$ T cells are the primary target cells of HIV-1, and progression towards AIDS is correlated with a decline in CD4$^+$ T cells. Nonetheless, there is accumulating evidence that CD4$^+$ T cells are important in the host response against HIV-1 infection, and the strength of the CD4$^+$ T-cell proliferative responses to HIV antigens is inversely correlated with the viral load. Moreover, patients that do not progress towards AIDS long after HIV-1 infection possess strong CD4$^+$ T-cell proliferative responses. Furthermore, recovery of CD4$^+$ T-cell proliferative responses to HIV antigens is related to the early treatment of acutely infected individuals with antiretroviral drugs (58).

CD4$^+$ T-cell-mediated help is controlled by MHC class II molecules, designated in humans as HLA-DP, -DQ and -DR, which all possess one *A* gene and at least one *B* gene. Recent publications show the importance of MHC class II-restricted CD4$^+$ T-helper cells in HIV-1 and SIV infection (59, 60). A study in a group of commercial sex workers in Nairobi, Kenya,

reported that particular *HLA-DRB* alleles are associated with resistance/susceptibility to HIV-1 infection (60). The study involving SIV-infected Indian rhesus macaques showed that elite control in these animals is correlated with the presence of certain rhesus macaque DRB molecules (59).

In humans, nine different *DRB* genes have been characterized (*HLA-DRB1* to *-DRB9*), and orthologs of these genes are present in chimpanzees (61). Humans and two macaque species have been haplotyped for their *DRB* region using one microsatellite marker, D6S2878 (62, 63). We used this marker to accurately haplotype the *DRB* region of our chimpanzee population (**Chapter 6**). This resulted in a revision of the number of region configurations present in the West-African chimpanzee population studied from six to seven. The configurations vary in gene content from two to five, and the different genes display none to moderate allelic variation. Humans, however, possess five main region configurations that contain one up to four *DRB* genes, and all region configurations display an abundant amount of allelic variation for their *DRB1* locus. Subsequently, we investigated which *Patr-DRB* alleles on a haplotype were transcribed, and could play a role in providing $CD4^+$ T-cell-mediated help to $CD8^+$ memory T cells (**Chapter 6**). Most haplotypes transcribe two *DRB* alleles, while one haplotype contained one, and five haplotypes contained three transcribed *DRB* alleles. In humans, only haplotypes with one or two transcribed *DRB* alleles are known, while rhesus macaque haplotypes transcribe one to three *DRB* alleles. These data suggest that three transcribed *DRB* alleles per haplotype is most likely the upper limit. This seems plausible, as the expression of too many MHC class II molecules would result in the deletion of a high number of T cells during thymic education.

A phylogenetic comparison of the transcribed *Patr-DRB* alleles with the *HLA-DRB* alleles that are associated with resistance/susceptibility to HIV-1 infection was not sufficient to define the contribution of *Patr-DRB* alleles in HIV-1/SIV_{cpz} infection. The fact that nine antigens of the HLA-B44 supertype possess unique peptide binding motifs (64), and the knowledge that MHC class II molecules have a promiscuous binding cleft (65, 66), further demonstrates that caution is needed when drawing conclusions based only on phylogenetic comparisons with regard to functional properties. Nevertheless, the presently available *Patr-DRB* alleles seem to be efficient in the maintenance of $CD4^+$ T-cell responses, and in such a way chimpanzees have successful combinations of $CD4^+$/$CD8^+$ T-cell responses that can cope with pathogens of their natural habitat and with HIV-1/SIV_{cpz} (**Chapter 6**).

The selective sweep in chimpanzees: Is an HIV-1/SIV_{cpz}-like retrovirus the causative agent?

We hypothesized that the selective sweep that affected the chimpanzee MHC class I repertoire was caused by HIV-1/SIV_{cpz} or a related ancestral retrovirus. MHC class I molecules play a pivotal role in the adaptive immune response, and present antigenic peptides of 8-10 amino acids in length to cytotoxic T cells (CTL).

In HIV-1-infected humans, adaptive immune responses to the Gag protein are considered to be important in the control of viral replication (67, 68). In particular, the MHC class I molecules HLA-B*2705 and -B*5701 are strongly associated with resistance to the development of AIDS in humans (13). To investigate the selective sweep hypothesis and whether this repertoire skewing resulted in the preferential selection of Patr molecules that are similar to the AIDS-resistant molecules in humans, we determined the peptide binding motifs of four Patr class I molecules (Patr-A*0301, -B*0101, -B*0301, and -B*0501) that occur at a high frequency in the wild-caught West-African chimpanzee population studied. The peptide binding motifs were used to scan the HIV-1 Gag protein for potential CTL epitopes, which were subsequently tested in peptide binding studies (**Chapter 7**). The results revealed that the four selected Patr molecules, like the AIDS-protective HLA-B*27 and -B*57 molecules, bind peptides derived from conserved areas of the HIV-1/SIV_{cpz} Gag-protein. 94% of the wild-caught West-African chimpanzees studied possess at least one of these four Patr class I molecules. Moreover, many chimpanzees were observed to express several molecules, all of which are able to bind peptides derived from various conserved regions; this suggests that chimpanzees may developed a "double-lock" strategy to respond to an HIV-1/SIV_{cpz} infection. A re-evaluation of immune response data with the gathered information revealed that three chimpanzees of an experimentally infected HIV-1 cohort displayed broadly reactive CTL responses. In conclusion, the present functional characteristics of the skewed chimpanzee MHC class I repertoire suggest that the ancient selective sweep was caused by a lentiviral pandemic. As a consequence most chimpanzees are able to cope with retroviral infections such as HIV-1/SIV_{cpz}, like AIDS-resistant HLA-B*27/B*57-positive humans (**Chapter 7**).

A recombinant virus emerging after a co-infection with three different HIV-1 strains provoked AIDS or AIDS-like symptoms in chimpanzees (69, 70). Particular chimpanzees (East-African subspecies) from a naturally SIV_{cpz}-infected cohort developed an AIDS-like immunopathology as well (71). Natural infections with SIV_{cpz} were observed in Central and East-African chimpanzee populations (7), and these SIV_{cpz} strains represent a mosaic of Old World monkey virus segments (72). Chimpanzees and their ancestors were most likely infected by predating on different monkey species that were infected with several types of SIV strains. It is plausible that new recombinant SIV_{cpz} strains are generated occasionally (73), and that some of these strains have pathogenic characteristics (**Chapter 7**). In general, naturally occurring SIV infections co-evolved with the host immune system, and lack progression towards AIDS. Some cases of progression to simian AIDS have been described in naturally or experimentally infected African green monkeys, Sooty mangabeys, and mandrills (74), which illustrates that SIV strains can be pathogenic. However, most hosts seem able to control the virus up to the time of their death from other causes (74).

Natural history has provided a few examples in which the introduction of a "novel" virus into a naïve population resulted in mass mortality. For instance, contact between European individuals and the native population of America led to the death of millions of Amerindians, as a result of measles, small pox, and other viral infections (75). The deliberate infection of rabbits with the myxoma virus, which exterminated approximately 99% of the rabbit

population in Australia, is another example (76). Natural infections in animals can have strong effects on population size, as is illustrated by the rinderpest epidemic in African buffalo (77), and the decimation of seals by morbilli and influenza viruses (78, 79).
The rhesus macaque, an Old World monkey species not naturally infected with SIV, is shown to develop AIDS, and most animals ultimately die after experimental infection with the virus (10-12). In contrast, the experimental introduction of HIV-1 in chimpanzees (a species later found to be naturally infected with SIV_{cpz}) resulted in general in the control of the viral infection. In addition, two naturally infected East-African chimpanzees that have been monitored carefully during captivity for more than 10 years did not develop any signs of AIDS (9). Moreover, AIDS did not develop after the experimental introduction of a natural SIV_{cpz} strain into animals of the West-African subspecies (9). These observations illustrate that the chimpanzee immune system has developed ways to control HIV-1/SIV_{cpz} infections (**Chapter 7**).

Conclusions and implications

This thesis describes findings with regard to the MHC class I repertoire of chimpanzees having experienced a selective sweep, which was most likely caused by a SIV-like virus. The resulting repertoire skewing favored the survival of animals that possessed Patr molecules, which target similar conserved areas of HIV-1/SIV_{cpz} as does HLA-B*27/B*57-positive human LTNP. As such, chimpanzees living at present possess MHC class I molecules that elicit CTL responses that have the capacity to control an HIV-1/SIV_{cpz} infection.
Modern humans, who have existed for about 150,000-200,000 years (52), contain MHC class I loci displaying an extensive amount of variation. However, without treatment, control of HIV-1 replication is most consistently associated with the presence of the HLA-B*27 and -B*57 molecules that target conserved areas of the HIV-1 Gag protein. The importance of the MHC region in controlling HIV-1 replication is strengthened by genome-wide association studies in different HIV-1-infected cohorts (80, 81). Recently, several reports stated that escape of the virus in the immunodominant regions of the Gag protein targeted by HLA-B*27 and -B*57 results in a loss of viral fitness (82-84). This suggest that an immune response that is directed to conserved areas of HIV-1 is a method to control viral replication.
In general, a virus survives best if it is able to replicate and disseminate itself within a population without killing its host. Such a host/virus state of equilibrium is, for instance, reached in non-human primates infected with SIV. For HIV-1 and its human hosts, the battle seems to be in full swing. Nevertheless, HIV-1 has reached a certain stage of equilibrium with HLA-B*27/B*57-positive individuals. At least these individuals can survive the infection for a long period in which reproduction of the host is possible. While other genetic systems probably contribute as well in controlling an HIV-1 infection, the findings described in this thesis suggest that without any intervention, survivors of the HIV-1 pandemic may be those individuals who are able to mount immune responses to conserved regions of the virus. These observations have important implications for vaccine design, since different HLA specificities

may target different conserved elements of HIV-1, which need to be identified. One major implication is that many individuals may not possess HLA-class I molecules that have the capacity to bind conserved HIV-1 epitopes. The promising results recently announced for the RV144 HIV-1 vaccine trial are raising hopes that a vaccine will be available in the not too distant future (www.hivresearch.org). Until then, however, the prevention of the infection itself will be, for the individual, one of the most important ways to survive the HIV-1 pandemic.

References

1. Barre-Sinoussi F, Chermann JC, Rey F, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.
2. Drew WL, Mintz L, Miner RC, et al. (1981) Prevalence of cytomegalovirus infection in homosexual men. J Infect Dis 143:188-192.
3. Levy JA (1993) Pathogenesis of human immunodeficiency virus infection. Microbiol Rev 57:183-289.
4. (2007) HIV vaccine failure prompts Merck to halt trial. Nature 449:390.
5. (2009) Back to basics. Nat Med 15:821.
6. Sodora DL, Allan JS, Apetrei C, et al. (2009) Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. Nat Med 15:861-865.
7. Santiago ML, Rodenburg CM, Kamenya S, et al. (2002) SIVcpz in wild chimpanzees. Science 295:465.
8. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287:607-614.
9. Heeney JL, Dalgleish AG, Weiss RA (2006) Origins of HIV and the evolution of resistance to AIDS. Science 313:462-466.
10. Daniel MD, Letvin NL, King NW, et al. (1985) Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. Science 228:1201-1204.
11. Kanki PJ, McLane MF, King NW, Jr., et al. (1985) Serologic identification and characterization of a macaque T-lymphotropic retrovirus closely related to HTLV-III. Science 228:1199-1201.
12. Letvin NL, Daniel MD, Sehgal PK, et al. (1985) Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III. Science 230:71-73.
13. Kiepiela P, Leslie AJ, Honeyborne I, et al. (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 432:769-775.
14. McAdam SN, Boyson JE, Liu X, et al. (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421-6429.
15. Connor RI, Sheridan KE, Ceradini D, et al. (1997) Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med 185:621-628.

16. van Kooyk Y, Geijtenbeek TB (2003) DC-SIGN: escape mechanism for pathogens. Nat Rev Immunol 3:697-709.
17. Liu H, Carrington M, Wang C, et al. (2006) Repeat-region polymorphisms in the gene for the dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related molecule: effects on HIV-1 susceptibility. J Infect Dis 193:698-702.
18. Zhang J, Zhang X, Fu J, et al. (2008) Protective role of DC-SIGN (CD209) neck-region alleles with <5 repeat units in HIV-1 transmission. J Infect Dis 198:68-71.
19. Geijtenbeek TB, Koopman G, van Duijnhoven GC, et al. (2001) Rhesus macaque and chimpanzee DC-SIGN act as HIV/SIV gp120 trans-receptors, similar to human DC-SIGN. Immunol Lett 79:101-107.
20. Bashirova AA, Wu L, Cheng J, et al. (2003) Novel member of the CD209 (DC-SIGN) gene family in primates. J Virol 77:217-227.
21. Santos PR, Michel-Salzat A, Butor C (2005) Chimpanzee DC-SIGN alleles predict the existence of A and B isoforms, but do not support a role for resistance to HIV infection. AIDS Res Hum Retroviruses 21:820-829.
22. Gagneux P, Wills C, Gerloff U, et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082.
23. Morin PA, Moore JJ, Chakraborty R, et al. (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201.
24. Hvilsom C, Carlsen F, Siegismund HR, et al. (2008) Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene. Genomics 92:322-328.
25. Gonzalez E, Bamshad M, Sato N, et al. (1999) Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. Proc Natl Acad Sci U S A 96:12004-12009.
26. Samson M, Libert F, Doranz BJ, et al. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382: 722-725.
27. Benton PA, Lee DR, Kennedy RC (1998) Sequence comparisons of non-human primate HIV-1 coreceptor homologues. Mol Immunol 35:95-101.
28. Pretet JL, Zerbib AC, Girard M, et al. (1997) Chimpanzee CXCR4 and CCR5 act as coreceptors for HIV type 1. AIDS Res Hum Retroviruses 13:1583-1587.
29. Voevodin A, Samilchuk E, Dashti S (1998) A survey for 32 nucleotide deletion in the CCR-5 chemokine receptor gene (deltaccr-5) conferring resistance to human immunodeficiency virus type 1 in different ethnic groups and in chimpanzees. J Med Virol 55:147-151.
30. McDermott DH, Colla JS, Kleeberger CA, et al. (2000) Genetic polymorphism in CX3CR1 and risk of HIV disease. Science 290:2031.
31. MacFie TS, Nerrienet E, de Groot NG, et al. (2009) Patterns of diversity in HIV-related loci among subspecies of chimpanzee: concordance at CCR5 and differences at CXCR4 and CX3CR1. Mol Biol Evol 26:719-727.
32. Liegeois F, Lafay B, Formenty P, et al. (2009) Full-length genome characterization of a novel simian immunodeficiency virus lineage (SIVolc) from olive Colobus (Procolobus verus) and new SIVwrcPbb strains from Western Red Colobus (Piliocolobus badius badius) from the Tai Forest in Ivory Coast. J Virol 83:428-439.
33. Wooding S, Stone AC, Dunn DM, et al. (2005) Contrasting effects of natural selection on human and chimpanzee CC chemokine receptor 5. Am J Hum Genet 76:291-301.

34. Mummidi S, Bamshad M, Ahuja SS, et al. (2000) Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. J Biol Chem 275:18946-18961.
35. Brenchley JM, Schacker TW, Ruff LE, et al. (2004) $CD4^+$ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J Exp Med 200:749-759.
36. Mehandru S, Poles MA, Tenner-Racz K, et al. (2004) Primary HIV-1 infection is associated with preferential depletion of $CD4^+$ T lymphocytes from effector sites in the gastrointestinal tract. J Exp Med 200:761-770.
37. Veazey RS, DeMaria M, Chalifoux LV, et al. (1998) Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science 280:427-431.
38. Veazey RS, Lackner AA (2004) Getting to the guts of HIV pathogenesis. J Exp Med 200: 697-700.
39. Groh V, Bahram S, Bauer S, et al. (1996) Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proc Natl Acad Sci U S A 93:12445-12450.
40. Bahram S (2000) MIC genes: from genetics to biology. Adv Immunol 76:1-60.
41. Robinson J, Waller MJ, Parham P, et al. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31: 311-314.
42. Steinle A, Groh V, Spies T (1998) Diversification, expression, and gamma delta T cell recognition of evolutionarily distant members of the MIC family of major histocompatibility complex class I-related molecules. Proc Natl Acad Sci U S A 95:12510-12515.
43. Cattley SK, Longman N, Dawkins RL, et al. (1999) Phylogenetic analysis of primate MIC (PERB11) sequences suggests that the representation of the gene family differs in different primates: comparison of MIC (PERB11) and C4. Eur J Immunogenet 26:233-238.
44. Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122.
45. Anzai T, Shiina T, Kimura N, et al. (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc Natl Acad Sci U S A 100:7708-7713.
46. Bauer S, Groh V, Wu J, et al. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285:727-729.
47. Gleimer M, Parham P (2003) Stress management: MHC class I and class I-like molecules as reporters of cellular stress. Immunity 19:469-477.
48. Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science 279:1737-1740.
49. Shum BP, Flodin LR, Muir DG, et al. (2002) Conservation and variation in human and common chimpanzee CD94 and NKG2 genes. J Immunol 168:240-252.
50. Fujiyama A, Watanabe H, Toyoda A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-134.
51. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713.

52. Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc Natl Acad Sci U S A 103:9381-9386.
53. Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.
54. Zhao Z, Jin L, Fu YX, et al. (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci U S A 97:11354-11358.
55. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
56. Webster MT, Smith NG, Ellegren H (2002) Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. Proc Natl Acad Sci U S A 99:8748-8753.
57. Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160:753-763.
58. Johnson WE, Desrosiers RC (2002) Viral persistence: HIV's strategies of immune system evasion. Annu Rev Med 53:499-518.
59. Giraldo-Vela JP, Rudersdorf R, Chung C, et al. (2008) The major histocompatibility complex class II alleles Mamu-DRB1*1003 and -DRB1*0306 are enriched in a cohort of simian immunodeficiency virus-infected rhesus macaque elite controllers. J Virol 82:859-870.
60. Lacap PA, Huntington JD, Luo M, et al. (2008) Associations of human leukocyte antigen DRB with resistance or susceptibility to HIV-1 infection in the Pumwani Sex Worker Cohort. Aids 22:1029-1038.
61. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167:339-350.
62. de Groot N, Doxiadis GG, de Vos-Rouweler AJ, et al. (2008) Comparative genetics of a highly divergent DRB microsatellite in different macaque species. Immunogenetics 60:737-748.
63. Doxiadis GG, de Groot N, Claas FH, et al. (2007) A highly divergent microsatellite facilitating fast and accurate *DRB* haplotyping in humans and rhesus macaques. Proc Natl Acad Sci U S A 104:8907-8912.
64. Hillen N, Mester G, Lemmel C, et al. (2008) Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. Eur J Immunol 38:2993-3003.
65. Hammer J, Valsasnini P, Tolba K, et al. (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. Cell 74:197-203.
66. Rammensee HG, Falk K, Rotzschke O (1993) MHC molecules as peptide receptors. Curr Opin Immunol 5:35-44.
67. Borghans JA, Molgaard A, de Boer RJ, Kesmir C (2007) HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. PLoS ONE 2:e920.
68. Goulder PJ, Watkins DI (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol 8:619-630.
69. Novembre FJ, de Rosayro J, Nidtha S, et al. (2001) Rapid CD4(+) T-cell loss induced by human immunodeficiency virus type 1(NC) in uninfected and previously infected chimpanzees. J Virol 75:1533-1539.
70. Novembre FJ, Saucier M, Anderson DC, et al. (1997) Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. J Virol 71:4086-4091.
71. Keele BF, Jones JH, Terio KA, et al. (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460:515-519.
72. Bailes E, Gao F, Bibollet-Ruche F, et al. (2003) Hybrid origin of SIV in chimpanzees. Science 300:1713.

73. Leitner T, Dazza MC, Ekwalanga M, et al. (2007) Sequence diversity among chimpanzee simian immunodeficiency viruses (SIVcpz) suggests that SIVcpzPts was derived from SIVcpzPtt through additional recombination events. AIDS Res Hum Retroviruses 23:1114-1118.
74. Pandrea I, Silvestri G, Apetrei C (2009) AIDS in african nonhuman primate hosts of SIVs: a new paradigm of SIV infection. Curr HIV Res 7:57-72.
75. McMichael AJ (2004) Environmental and social influences on emerging infectious diseases: past, present and future. Philos Trans R Soc Lond B Biol Sci 359:1049-1058.
76. Silvers L, Inglis B, Labudovic A, et al. (2006) Virulence and pathogenesis of the MSW and MSD strains of Californian myxoma virus in European rabbits with genetic resistance to myxomatosis compared to rabbits with no genetic resistance. Virology 348:72-83.
77. Wenink PW, Groen AF, Roelke-Parker ME, Prins HH (1998) African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. Mol Ecol 7:1315-1322.
78. Domingo M, Visa J, Pumarola M, et al. (1992) Pathologic and immunocytochemical studies of morbillivirus infection in striped dolphins (Stenella coeruleoalba). Vet Pathol 29:1-10.
79. Geraci JR, St Aubin DJ, Barker IK, et al. (1982) Mass mortality of harbor seals: pneumonia associated with influenza A virus. Science 215:1129-1131.
80. Dalmasso C, Carpentier W, Meyer L, et al. (2008) Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS ONE 3:e3907.
81. Fellay J, Shianna KV, Ge D, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317:944-947.
82. Miura T, Brockman MA, Schneidewind A, et al. (2009) HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. J Virol 83:2743-2755.
83. Schneidewind A, Brockman MA, Yang R, et al. (2007) Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J Virol 81:12382-12393.
84. Troyer RM, McNevin J, Liu Y, et al. (2009) Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. PLoS Pathog 5:e1000365.

Abbreviations

AIDS	acquired immunodeficiency syndrome
AR	antiretroviral
BPRC	Biomedical Primate Research Centre
cDNA	complementary deoxyribonucleic acid
CI	confidence interval
CPBCA	cell-based peptide binding competition assay
CTL	cytotoxic T lymphocyte
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
FCS	fetal calf serum
gDNA	genomic deoxyribonucleic acid
Gogo	*Gorilla gorilla*
HBV	hepatitis B virus
HCV	hepatitis C virus
H_E	heterozygosity index
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPLC	high performance liquid chromatographic
IMDM	iscove's modified dulbecco's media
kb	kilo base pairs
KIR	Killer-cell immunoglobulin-like receptor
Ma	million years ago
Mamu	*Macaca mulatta*
MHC	major histocompatibility complex
MIC	major histocompatibility chain-related gene
mtDNA	mitochondrial deoxyribonucleic acid
MS	mass spectrometry
MYA	million years ago
myr	million years
N_0	current population size
N_I	ancestral population size
n_e	number of unique alleles
NJ	neighbor-joining
NK cell	natural killer cell
NSI	non-syncytium inducing
Papa	*Pan paniscus*
Patr	*Pan troglodytes*
PBS	phosphate buffered saline

PCR	polymerase chain reaction
Popy	*Pongo pygmaeus*
P.t.t.	*Pan troglodytes troglodytes*
P.t.s.	*Pan troglodytes schweinfurthii*
P.t.v.	*Pan troglodytes verus*
RNA	ribonucleic acid
RP-HPLC	reverse phase-high performance liquid chromatographic
RSCA	reference strand conformation analysis
RT	reverse transcriptase
SAL	single molecule-expressing cell-line
SC	standardized competitor
SI	syncytium inducing
SIV_{cpz}	chimpanzee-derived simian immunodeficiency virus
STR	short tandem repeat
t_{fa}	time in years since the population declined or expanded

Summary

Chimpanzees are, from an evolutionary perspective, humans closest living relatives. They are susceptible to infection with human immunodeficiency virus type-1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS). However, most chimpanzees are relative resistant to develop AIDS-mediated pathology after HIV-1 infection. Particular human individuals possess natural immunity to HIV-1 infection as well, the so-called long-term non-progressors (LTNP). Whether chimpanzees and human LTNP share certain genetic features was investigated.

The MHC, located on chromosome 6, plays a central role in the immune system, and has a key function in controlling immune responses to viruses. It can be divided into a class I and II region. The class I region encodes molecules that present peptides from intracellular origin to cytolytic $CD8^+$ T cells, whereas the class II region encodes molecules that present peptides from extracellular origin, and controls antibody responses and $CD4^+$ T cell mediated help. In humans and chimpanzees the molecules are designated HLA- and Patr-A, -B, -C, and -DP, -DQ, -DR, respectively. The common ancestry of both species is reflected in the sharing of genes/loci, but also particular lineages predate their speciation.

This thesis starts with the detailed characterization of the *Mhc* class I repertoire of a pedigreed West-African chimpanzee colony (**Chapter 2**). This survey showed that chimpanzees have only one of the six human *A* locus lineages, which suggested some kind of natural selection. In addition, the analysis revealed that within the small population of animals analyzed also the *B* and *C* loci showed signs of a repertoire reduction. Within the relatively few lineages detected a rather abundant allelic polymorphism was observed. However, *Mhc* class I and II sequences encode for molecules that experience frequency-dependent/diversifying selection. Furthermore, the imbalance in sample size between the number of humans and chimpanzees analyzed could hamper the correct interpretation of the data. As such, caution is necessary to answer the question if the chimpanzee MHC repertoire was skewed due to selection. Therefore the potential influence of negative/purifying selection operating on the *Mhc* class I genes was studied by analyzing the intron variation. This evidenced that chimpanzees indeed experienced a selective sweep that was most prominent at the *Patr-B* locus (**Chapter 3**). The hypothesis was put forward that the selective sweep was caused by a chimpanzee-derived simian immunodeficiency virus (SIV_{cpz}) or a closely related retrovirus, implying that the contemporary chimpanzee populations represent the off-spring of AIDS-resistant animals, the survivors of a HIV-1-like pandemic that took place in the distant past (**Chapter 3**).

To investigate whether the selective sweep affected other loci located on chromosome 6, the major histocompatibility complex class I chain related gene (*MIC*), mapping next to the *B* locus was studied (**Chapter 4**). In humans a *MICA* and *B* gene are found, which both display polymorphism. In contrast, chimpanzees have only one *MIC* gene that appeared to originate from a fusion between *MICA* and *B*, controlling only one lineage,

showing moderate allelic variation. Hence, the reduced *MIC* repertoire is in agreement with the selective sweep observed for the *Patr-B* locus. In addition, a multilocus comparison between humans and chimpanzees using microsatellite markers located in the MHC region, and markers mapping at a variety of other chromosomes revealed that both species show similar amounts of variation for the non-MHC markers, but in chimpanzees a significant lower allelic variation for most MHC markers was observed (**Chapter 5**). This confirmed again that chimpanzees possess a reduced MHC repertoire. Multilocus demographic analyses underscored that chimpanzees experienced a selective sweep that mainly targeted the chromosomal segment carrying the MHC class I region, and that other loci, such as the MHC class II loci and MIC, are probably affected due to genetic linkage. Subsequent analysis showed that most chimpanzees possess unique MHC haplotypes, suggesting that naturally occurring recombination-like processes allowed the establishment of haplotype diversity after the selective sweep (**Chapter 5**).

The accurate haplotyping of the *Patr-DRB* region with the use of a microsatellite marker resulted in the definition of seven region configurations (**Chapter 6**). The configurations vary in gene content from two to five, of which one up to three were found to be transcribed. The limited allelic variation observed, however, is in agreement with a selective sweep targeting the MHC region.

Finally we whished to examine the selective sweep hypothesis at the functional level, and whether it relates to AIDS-resistance as observed in HLA-B*27/B*57-positive human LTNP. Therefore it was investigated whether different Patr-A and -B molecules are able to present conserved HIV-1/SIV_{cpz} Gag peptides, which contribute in controlling viral replication (**Chapter 7**). The four selected Patr molecules were, like the AIDS-protective HLA-B*27 and -B*57, able to bind peptides derived from conserved areas of the HIV-1/SIV_{cpz} Gag-protein. Besides, sharing of binding motifs was observed. 94% of the chimpanzees studied possessed at least one Patr class I molecule that has the characteristics of a human molecule associated with AIDS resistance. In addition, many chimpanzees appear to express several molecules able to bind multiple peptides derived from various conserved Gag regions, suggesting that chimpanzees have developed a "double-lock" strategy.

In conclusion, this thesis describes that the MHC class I repertoire of chimpanzees has been skewed due to a selective sweep favoring the survival of animals that have Patr molecules targeting similar conserved areas of HIV-1/SIV_{cpz} as does HLA-B*27/B*57-positive human LTNP.

Samenvatting

Chimpansees, evolutionair gezien nauw verwant aan de mens, zijn, net als de mens, gevoelig voor infecties met het "human immunodeficiency virus type-1" (HIV-1); het virus dat acquired immuno deficiency syndrome (AIDS) veroorzaakt. Echter, de meeste HIV-1 geïnfecteerde chimpansees ontwikkelen geen AIDS verschijnselen. Er zijn ook mensen die een natuurlijke immuniteit tegen HIV-1 bezitten. Deze specifieke groep wordt aangeduid als "long-term non-progressors" (LTNP). Of chimpansees en de LTNP bepaalde genetische eigenschappen gemeen hebben werd in dit proefschrift onderzocht.

Het MHC, gelokaliseerd op chromosoom 6, speelt een centrale rol in het immuunsysteem, en heeft een sleutelfunctie in het aansturen van immunologische reacties tegen virussen. Het is verdeeld in een klasse I en II regio. De klasse I regio codeert voor moleculen die intracellulair gedegradeerde eiwitten (peptiden) presenteren aan $CD8^+$ T cellen. De klasse II regio codeert voor moleculen die peptiden presenteren van extracellulaire origine, en zorgen voor antilichaamproductie en $CD4^+$ T cel gemedieerde help. In de mens en chimpansee worden de moleculen respectievelijk HLA en Patr-A, -B, -C en -DP, -DQ, -DR genoemd. De verwantschap tussen mens en chimpansee weerspiegeld zich niet alleen in het delen van gelijke genen, maar ook in het delen van bepaalde groepen van nauw verwante allelen (lineages). Dit impliceert dat dergelijke genen en groepen al aanwezig waren voor de speciatie.

Dit proefschrift begint met de karakterisatie van het *Mhc* klasse I repertoire van een West-Afrikaanse chimpansee populatie (**Hoofdstuk 2**). Deze analyse toonde aan dat chimpansees maar één van de zes humane *Mhc* klasse I *A* "lineages" hebben, wat de suggestie wekte dat chimpansees een natuurlijke selectie hebben ondergaan. Tevens toonde de studie aan dat de *B* en *C* genen mogelijk ook een repertoire reductie hebben ondergaan. Maar binnen de beperkte hoeveelheid aanwezige "lineages" werd wel een aanzienlijke hoeveelheid allelische diversiteit gevonden. Verder is bekend dat *Mhc* klasse I en II sequenties coderen voor moleculen die onder positieve selectie druk staan. Tevens kan het verschil in het aantal geanalyseerde mensen en chimpansees de correcte interpretatie van de data bemoeilijken. Deze observaties tesamen zorgde ervoor dat we uit deze verkregen data niet zomaar konden concluderen dat chimpansees een natuurlijke selectie hebben ondergaan dat resulteerde in een beperkt *Mhc* repertoire.

Om de mogelijke invloed van selectie op het *Mhc* klasse I repertoire te bestuderen werd vervolgens de intron variatie geanalyseerd, omdat van intronen bekend is dat deze over het algemeen geen selectie druk ondervinden. Deze data toonde aan dat chimpansees wel degelijk een selectie hebben ondergaan, en dat deze het sterkst gericht was tegen het repertoire van het *Patr-B* gen (**Hoofdstuk 3**). Vervolgens is door ons gepostuleerd dat de repertoire reductie mogelijk veroorzaakt zou kunnen zijn door het simian immunodeficiency virus van de chimpansee (SIV_{cpz}) of een sterk gelijkend retrovirus. Als gevolg van deze natuurlijke selectie zijn de nu nog levende chimpansee populaties de nakomelingen van AIDS-resistente dieren, de overlevers van een HIV-1-achtige pandemie die in het verleden

heeft plaatsgevonden (**Hoofdstuk 3**).
Om te onderzoeken of de selectie ook van invloed is geweest op andere genen op chromosoom 6 werd het, naast het *B* gelegen, major histocompatibility complex klasse I gerelateerde gen (*MIC*) bestudeerd (**Hoofdstuk 4**). In de mens onderscheidt men een *MICA* en *B* gen, die beide polymorf zijn. Chimpansees, daarentegen, hebben maar één *MIC* gen dat een fusieproduct is tussen *MICA* en *B*. De studie toonde aan dat het *Patr-MICA/B* gen slechts één lineage kent, met een gemiddeld polymorf karakter. Dit gereduceerde *MIC* repertoire stemt overeen met de repertoire reductie die is gevonden voor het *Patr-B* gen.
Opvolgend werd de genomische diversiteit tussen mens en chimpansee vergeleken met behulp van microsatellietmerkers, die gelokaliseerd zijn in de MHC regio en op verschillende andere chromosomen (**Hoofdstuk 5**). Voor beide soorten werd, zoals verwacht, een gelijke hoeveelheid aan allelische variatie voor de niet-MHC merkers gevonden. Echter, chimpansees vertonen minder allelische variatie voor de meeste MHC merkers. Deze studie bevestigde dat de repertoire selectie vooral gericht was tegen dat gedeelte van chromosoom 6 waar de MHC klasse I regio ligt, en dat andere polymorfe genen, zoals de *Mhc* klasse II genen en *MIC*, beïnvloed werden door hun genetische koppeling. De observatie werd onderbouwd door middel van multilocus demografische analyses. Een aansluitende studie waarin de MHC klasse I en II haplotypen (combinatie van genen op een chromosoom) bepaald werden maakte duidelijk dat de meeste chimpansees unieke combinaties van *Mhc* allelen hebben, en suggereert dat na de repertoire selectie recombinatie-gerelateerde processen verantwoordelijk zijn geweest voor het genereren van haplotypen diversiteit (**Hoofdstuk 5**).
De haplotypering van de *Patr-DRB* regio met behulp van één microsatelietmerker resulteerde in de beschrijving van zeven verschillende *DRB* regio configuraties in de West-Afrikaanse chimpansee populatie (**Hoofdstuk 6**). De configuraties variëren in het aantal *DRB* genen van 2 tot 5, waarvan slechts 1 tot 3 genen worden getranscribeerd. De beperkte aanwezige allelische variatie stemt overeen met een repertoire selectie.
Het proefschrift eindigt met een hoofdstuk waarin de hypothese betreffende de repertoire reductie in chimpansees op het functionele vlak werd onderzocht. Er werd bestudeerd of er een verband bestaat tussen de AIDS-resistentie in chimpansees en de verhoogde aanwezigheid van HLA-B*27/B*57-positieve individuen in LTNP (**Hoofdstuk 7**). Van verschillende Patr-A and -B moleculen werd onderzocht of zij de capaciteit hebben om geconserveerde HIV-1/SIV_{cpz} Gag peptiden te binden. De vier geselecteerde Patr moleculen kunnen, net als de AIDS-resistente HLA-B*27/B*57 moleculen, peptide binden van geconserveerde delen van het HIV-1/SIV_{cpz} Gag eiwit. Bovendien bezitten sommige chimpansee moleculen overeenkomstige bindingsmotieven met HLA-B*27/B*57. Vierennegentig procent van de bestudeerde chimpansees heeft ten minste één Patr klasse I molecuul met de karakteristieken van een molecuul dat in de mens geassocieerd is met AIDS resistentie. Verder bezitten veel chimpansees meerdere MHC moleculen die de capaciteit hebben om verschillende peptide uit verscheidene geconserveerde gebieden van Gag te binden. Dit suggereerd dat chimpanseeseen "double-lock strategy" hebben ontwikkeld.

Tot slot, in dit proefschrift wordt beschreven dat het MHC klasse I repertoire van chimpansees een repertoire reductie heeft ondergaan. In het wild kunnen chimpansees geïnfecteerd raken met verschillende SIV stammen. De repertoire selectie was gunstig voor het voortbestaan van de dieren die Patr moleculen hebben die geconserveerde delen van HIV-1/SIV_{cpz} kunnen herkennen en presenteren net zoals dat het geval is voor HLA-B*27/B*57-positieve individuen in de menselijke populatie.

Dankwoord

Wat is er veel gebeurd de afgelopen jaren, maar het resultaat is er nu. Het proefschrift heeft zijn vorm gekregen. Dat ik dat uiteindelijk niet allemaal alleen voor elkaar heb gekregen mag duidelijk zijn. Het is dan ook fijn nu een dankwoord te mogen schrijven. Toch zullen de twee pagina's die ik hiervoor gereserveerd heb waarschijnlijk niet genoeg zijn. Daarom wil ik beginnen met iedereen (familie, vrienden, collega's) die de afgelopen jaren in meer of mindere mate in mijn naaste omgeving aanwezig waren met raad, steun, voor discussies, opvang van mijn kinderen etc. hierbij bedanken. Jullie aanwezigheid is van onschatbare waarde geweest en heeft mij gebracht waar ik nu ben.

Dan een persoonlijk woord voor jou Ronald. Bedankt dat je mij de mogelijkheid hebt gegeven mij als wetenschapper te kunnen ontplooien. Er is veel gebeurd in de tijd dat ik aan dit bijzondere onderwerp heb gewerkt, maar je hebt mij altijd gesteund en daar ben ik je ontzettend dankbaar voor. Ik hoop dat we nog lang samen mogen werken. En dan jij Su, want dertien jaar naar dato, maar hier is mijn boekje dan. Bedankt dat je jou vertrouwen in mijn kunnen hebt uitgesproken, je was een bijzonder mens. Inmiddels heeft jou beslissing om planeet aarde te verlaten een plekje gekregen, al blijft het altijd onbegrijpelijk.

Nel, bedankt dat je mij in de begin jaren de kneepjes van het vak hebt geleerd en altijd bereid was mijn manuscripten te lezen, je bent een fijne collega. Gaby, bedankt voor al die momenten dat ik bij je binnen kon lopen en alles wat me dwars zat met jou kon bespreken, op wetenschappenlijk en persoonlijk vlak. Daar heb ik enorm veel van geleerd, en ben blij dat je een van mijn paranimfen wilt zijn. Corrine, we werken al weer wat jaren samen. Eerst deelde we alleen een zitkamer, vanaf 2005 ben jij mij komen ondersteunen in mijn werkzaamheden. We hebben heel wat werk verzet sinds die tijd, naast dat je mijn tweede paar handen was dacht je ook altijd mee over de werkzaamheden. Bedankt, zonder jou was het werk nu nog niet klaar geweest. Ik hoop dat onze samenwerking zich nog lang voort kan zetten. Nanine, nog een tweede paar handen. Bedankt voor alle STR data die jij voor mij gegenereerd hebt, en waaruit twee mooie publicaties zijn gekomen. Met het opnoemen van Annemiek, Marit en Marije, heb ik alle babes van de afdeling CG&R gehad. We hebben een fijne groep dames en ik hoop dat we nog lang zo verder kunnen werken, op naar wetenschappenlijke succesjes. Ja, Jeroen, ik ben jou niet vergeten hoor, maar als "haan" in het "kippenhok" verdien je toch nog wat extra regels tekst. Het was een voorrecht om de afgelopen jaren met jou een kamer te mogen delen. Bedankt voor alle discussies en ideetjes. Je bent een enorme kennis bron en ik hoop dat je de afdeling nog lang blijft versterken. Ernst, Gerrit, Jeffrey, David, Jon, bedankt voor jullie kritische blik op mijn manuscripten en de mogelijkheid vragen te stellen en te discussiëren. Frank en Ed, bedankt met de hulp met de peptide-binding studies en statistiek. Babs, helaas heeft onze samenwerking tot nu toe nog geen resultaat gebracht, maar we geven niet op. Ik heb het afgelopen jaar wel veel geleerd, bedankt daarvoor. Henk, bedankt voor je geduld met het maken van alle fraaie figuren passend bij elk manuscript. Daarbij niet vergetend je humoristische plaatjes die je af en toe creëert en daarmee de boel in een ander

perspectief zet. Donna, thanks to you for all the intensive work of editing my manuscripts, you are truly invaluable.
Hierbij wil ik ook mijn dank uitspreken naar de leden van de beoordelingscommissie. Prof. dr. R.J. de Boer, Prof. dr. W. van Eden, Prof. dr. P. Hogeweg en Prof. dr. F. Miedema bedankt dat u bereid was mijn proefschrift te beoordelen. Prof. dr. D.I. Watkins, thanks that you excepted to be one of the members of the committee and that you will be present at the day of the defense.
Naast collega's, zijn familie en vrienden gedurende de afgelopen jaren ook van onschatbare waarde geweest. Jörgen en Patricia, wat fijn dat we zo een goede familieband hebben en zoveel gezellige dingen samen doen die mij dan even het werk doen laten vergeten. Ome Nico, bedankt voor u bijdrage aan dit proefschrift. Het is met recht een voorkant met een verhaal geworden, precies zoals ik gehoopt had. Ook jou Cin wil ik even bij toenaam noemen, je was er precies op het moment dat ik je het hardst nodig had. Het is een gemis dat we niet meer zo dicht bij elkaar wonen. Fijn dat je mijn paranimf wilt zijn zodat we deze bijzondere dag samen kunnen meemaken.
Pa en Ma, wie en wat zou ik zijn zonder jullie. Jullie hebben altijd in mij geloofd, hebben me overal in gesteund, staan altijd voor mij klaar. Mijn dank naar jullie is zo groot, dat het gewoon onbeschrijfbaar is, ik hoop dat jullie dat beseffen. Ik hoop dat we nog lang van elkaar mogen genieten.
Patrick, toen ik jou in '99 ontmoeten, had ik net tegen Bontrop gezegd dat ik wel promotie onderzoek wilde doen. Tussen toen en nu is er veel gebeurd. Inmiddels zijn we niet meer met zijn tweeën, maar vormen we met Mirte en Tycho een gezin. Met dit schrijven is er dan ook echt een eind gekomen aan mijn uitspraak "nou nog twee jaar denk ik en dan is het echt klaar". Bedankt dat je me altijd gesteund hebt. Je bent een schat en ik hou ontzettend veel van je. Ik hoop dat de toekomst ons nog veel mooie dingen zal brengen.

List of publications

1. G.G.M. Doxiadis, N. Otting, S.M.G. Antunes, N.G. de Groot, M. Harvey, I.I.N Doxiadis, M. Jonker, R.E. Bontrop. Characterization of the ABO blood group genes in macaques: evidence for convergent evolution. Tissue Antigens 51, 321-326, 1998.
2. N.Otting, G.G.M. Doxiadis, L. Versluis, N.G. de Groot, J. Anholts, W. Verduin, E. Rozemuller, F. Claas, M.G.J. Tilanus, R.E. Bontrop. Characterisation and distribution of Mhc-DPB1 alleles in chimpanzee and rhesus macaque populations. Human Immunology 59, 656-664, 1998.
3. S.G. Antunes, N.G. de Groot, H. Brok. G. Doxiadis, A.A.L. Menezes, N. Otting, R.E. Bontrop. The Common Marmoset: a New World primate species with limited Mhc class II variability. Proc. Natl. Acad. Sci. USA 59, 11745-11750, September 1998.
4. N.G. de Groot, N. Otting, G.G.M. Doxiadis, S.M.G. Antunes, R.E. Bontrop. Characterisation of four non-human primate MHC-DQB1 alleles. Tissue Antigens 52, 497-499, 1998.
5. R.E. Bontop, N. Otting, N.G. de Groot, G.G.M. Doxiadis. Major histocompatibility complex class II polymorphisms in primates. Immunological reviews 167, 339-350, 1999.
6. N.G. de Groot, R.E. Bontrop. The major histocompatibility complex class II region of the chimpanzee: towards a molecular map. Immunogenetics 50, 160-167, 1999.
7. G.G.M. Doxiadis, N. Otting, N.G. de Groot, R. Noort, R.E. Bontrop. Unprecedented polymorphism of MHC-DRB region configurations in rhesus macaques. Journal of Immunology 164, 3193-3199, 2000.
8. N.G. de Groot, N. Otting, R. Arguello, D.I Watkins, G.G.M. Doxiadis, J.A. Madrigal, R.E. Bontrop. MHC class I diversity in a West-African chimpanzee population: implications for HIV research, Immunogenetics 51, 398-409, 2000.
9. N.Otting, N.G. de Groot, M.C. Noort, G.G.M. Doxiadis, R.E. Bontrop. Allelic diversity of MHC-DRB alleles in rhesus macaques, tissue antigens 56, 58-68, 2000.
10. N. Otting, N.G. de Groot, M.C. Noort, G.G.M. Doxiadis, R.E. Bontrop. Allelic diversity of MHC-DRB alleles in rhesus macaques, tissue antigens 56, 58-68, 2000.
11. G.G.M. Doxiadis, N. Otting, N.G. de Groot, R.E. Bontrop. Differential evolutionary MHC class II strategies in humans and rhesus macaques: relevance for biomedical studies. Immunological reviews 183, 76-85, 2001
12. N. Otting, N.G de Groot, G.G.M. Doxiadis, R.E. Bontrop. Extensive Mhc-DQB variation in humans and non-human primate species, Immunogenetics 54, 230-239, 2002.
13. N.G. de Groot, N. Otting, G.G.M. Doxiadis, S.S. Balla-jhagjhoorsingh, J.L. Heeney, J.J van Rood, P. Gagneux, R.E. Bontrop. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees, Proc. Natl. Acad. Sci. USA 99, 11748-11753, 2002.
14. S.S. Balla-Jhagjhoorsingh, E.J. Verschoor, N.G. de Groot, V.J.P. Teeuwsen, R.E. Bontrop, J.L. Heeney. Specific nature of cellular immune response elicited by chimpanzees against HIV-1, Human Immunology 64, 681-688, 2003.
15. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SG. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex, Nucleic Acids Res. 31, 311-314, 2003.
16. Doxiadis GG, Otting N, de Groot NG, de Groot N, Rouweler AJ, Noort R, Verschoor EJ, Bontjer I, Bontrop RE. Evolutionary stability of MHC class I haplotypes in diverse rhesus macaque populations, Immunogenetics 55, 540-551, 2003.
17. de Groot N, Doxiadis GG, de Groot NG, Otting N, Heijmans C, Rouweler AJ, Bontrop RE. Genetic makeup of the DR region in rhesus macaques: gene content, transcripts, and pseudogenes, Journal of Immunology 172, 6152-6157, 2004.

18. Voorter CE, de Groot NG, Meertens CM, Bontrop RE, van den Berg-Loonen EM. Allelic polymorphism in introns 1 and 2 of the HLA-DQA1 gene, Tissue Antigens 65, 56-66, 2005.
19. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE. Unparalleled complexity of the MHC class I region in rhesus macaques, Proc. Natl. Acad. Sci. USA 102, 1626-1631, 2005.
20. de Groot NG, Garcia CA, Verschoor EJ, Doxiadis GGM, Marsh SGE, Otting N, Bontrop RE. Reduced MIC gene repertoire variation in West African chimpanzees as compared to humans. Mol. Biol. Evol. 22, 1375-1385, 2005.
21. Penedo MC, Bontrop RE, Heijmans CM, Otting N, Noort R, Rouweler AJ, de Groot N, de Groot NG, Ward T, Doxiadis GG. Microsatellite typing of the rhesus macaque MHC region, Immunogenetics 57, 198-209, 2005.
22. Doxiadis GG, Rouweler AJ, de Groot NG, Louwerse A, Otting N, Verschoor EJ, Bontrop RE. Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques, Immunogenetics 58, 259-68, 2006.
23. Doxiadis GG, van der Wiel MK, Brok HP, de Groot NG, Otting N, 't Hart BA, van Rood JJ, Bontrop RE. Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species, Proc. Natl. Acad. Sci. USA 103, 5864-8, 2006.
24. Otting N, de Vos-Rouweler AJ, Heijmans CM, de Groot NG, Doxiadis GG, Bonrop RE. MHC class I a region diversity and polymorphism in macaque species, Immunogenetics 59 (5), 367-375, 2007.
25. de Groot NG, Heijmans CM, de Groot N, Otting N, de Vos-Rouweler AJ, Bonhomme M, Doxiadis GG, Crouau-Roy B, Bontrop RE. Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics, Molecular Ecology 17(8), 2074-88, 2008.
26. Doxiadis GG, de Groot N, de Groot NG, Doxiadis I, Bontrop RE. Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: old wine served in new skins. Molecular Immunology 45(10), 2743-51, 2008.
27. Otting N, Heijmans CM, van der Wiel M, de Groot NG, Doxiadis GG, Bontrop RE. A snapshot of the Mamu-B genes and their allelic repertoire in rhesus macaques of Chinese origin. Immunogenetics 60(9), 507-14, 2008.
28. de Groot N, Doxiadis GG, de Vos-Rouweler AJ, de Groot NG, Verschoor EJ, Bontrop RE. Comparative genetics of a highly divergent DRB microsatellite in different macaque species. Immunogenetics 60(12), 737-48, 2008.
29. MacFie TS, Nerrienet E, de Groot NG, Bontrop RE, Mundy NI. Patterns of diversity in HIV-related loci among subspecies of chimpanzee: concordance at CCR5 and differences at CXCR4 and CX3CR1. Mol Biol Evol. 26(4), 719-27, 2009.
30. Garamszegi LZ, de Groot NG, Bontrop RE. Correlated evolution of nucleotide substitution rates and allelic variation in Mhc-DRB lineages of primates. BMC Evol Biol. 12, 73, 2009.
31. de Groot NG, Heijmans CM, de Groot N, Doxiadis GG, Otting N, Bontrop RE. The chimpanzee Mhc-DRB region revisited: gene content, polymorphism, pseudogenes, and transcripts. Molecular Immunology 47 (2-3), 381-89, 2009.
32. Doxiadis GG, de Groot N, de Groot NG, Rotmans G, de Vos-Rouweler AJ, Bontrop RE. Extensive DRB region diversity in cynomolgus macaques: recombination as a driving force. Immunogenetics 62 (3), 137-47, 2010.

Curriculum Vitae

De schrijfster van dit proefschrift werd geboren op 18 april 1972 in Den Haag. Op 15 Juni 1988 behaalde zij haar MAVO diploma aan het Aloysius College in Den Haag, waarna zij begon aan een MLO opleiding aan het Van Leeuwenhoek Instituut/Reynevelt College in Delft. Het diploma met als specialisatie Technische Microbiologie werd op 15 Juni 1992 behaald. De studie werd voortgezet aan de Internationale Agrarische Hogeschool Larenstein, afdeling laboratoriumtechniek; het eerste jaar in Wageningen en daarna in Velp. In de laatste fase van deze studie liep zij van Augustus 1995 tot en met December 1995 stage bij de afdeling Botany verbonden aan de Universiteit van Britisch Columbia in Vancouver, Canada. Vervolgens werkte zij van Januari 1996 tot en met Mei 1996 aan een afstudeerproject bij de afdeling Immunobiologie van het Biomedical Primate Research Centre (BPRC), waarna op 19 Juni 1996 de opleiding met als afstudeerrichting Microbiologie/Biochemie met succes werd afgerond. Op 1 Augustus 1996 trad zij als research medewerker in dienst van het BPRC. In September 1999 start zij met haar promotie onderzoek dat werd uitgevoerd bij de afdeling Comparative Genetics and Refinement van het BPRC.

The author of this thesis was born on April 18, 1972 in The Hague, The Netherlands. On June 15, 1988 she obtained her MAVO diploma (Aloysius College, The Hague), after which she started a MLO study at the Van Leeuwenhoek Institute/Reynevelt College in Delft. The diploma with a specialization in Technical Microbiology was achieved on June 15, 1992. The study continued at Larenstein International Agricultural College, Department laboratory technique; the first year in Wageningen and then in Velp. In the final part of this study she accomplished a practical training period from August 1995 till December 1995 at the department of Botany at the University of Britisch Columbia in Vancouver, Canada. From January 1996 to May 1996 she accomplished a graduate assignment at the Department of Immunobiology at the Biomedical Primate Research Centre (BPRC). On June 19, 1996 the study with a specialization in Microbiology/Biochemistry was completed successfully. On August 1, 1996, she was employed as a research assistant at the BPRC. In September 1999 she started her PhD project that was performed at the Department of Comparative Genetics and Refinement of the BPRC.