

# Selective Pressures of HLA Genotypes and Antiviral Therapy on Human Immunodeficiency Virus Type 1 Sequence Mutation at a Population Level<sup>∇</sup>

Golo Ahlenstiel,<sup>1†</sup> Kirsten Roomp,<sup>2†</sup> Martin Däumer,<sup>3</sup> Jacob Nattermann,<sup>1</sup> Martin Vogel,<sup>1</sup> Jürgen K. Rockstroh,<sup>1</sup> Niko Beerenwinkel,<sup>4</sup> Rolf Kaiser,<sup>3</sup> Hans-Dieter Nischalke,<sup>1</sup> Tilman Sauerbruch,<sup>1</sup> Thomas Lengauer,<sup>2</sup> and Ulrich Spengler<sup>1\*</sup>  
 on behalf of the Kompetenznetz HIV/AIDS

*Department of Internal Medicine I, University of Bonn, 53105 Bonn, Germany<sup>1</sup>; Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, 66123 Saarbrücken, Germany<sup>2</sup>; Institute of Virology, University of Cologne, 50935 Cologne, Germany<sup>3</sup>; and Department of Mathematics, University of California, Berkeley, California<sup>4</sup>*

Received 18 April 2007/Returned for modification 31 May 2007/Accepted 2 August 2007

**The objective of this study was a comprehensive analysis of the immune-driven evolution of viruses of human immunodeficiency virus type 1 (HIV-1) clade B in a large patient cohort treated at a single hospital in Germany and its implications for antiretroviral therapy. We examined the association of the HLA-A, HLA-B, and HLA-DRB1 alleles with the emergence of mutations in the complete protease gene and the first 330 codons of the reverse transcriptase (RT) gene of HIV-1, studying their distribution and persistence and their impact on antiviral drug therapy. The clinical data for 179 HIV-infected patients, the results of HLA genotyping, and virus sequences were analyzed using a variety of statistical approaches. We describe new HLA-associated mutations in both viral protease and RT, several of which are associated with HLA-DRB1. The mutations reported are remarkably persistent within our cohort, developing more slowly in a minority of patients. Interestingly, several HLA-associated mutations occur at the same positions as drug resistance mutations in patient viruses, where the viral sequence was acquired before exposure to these drugs. The influence of HLA on thymidine analogue mutation pathways was not observed. We were able to confirm immune-driven selection pressure by major histocompatibility complex (MHC) class I and II alleles through the identification of HLA-associated mutations. HLA-B alleles were involved in more associations (68%) than either HLA-A (23%) or HLA-DRB1 (9%). As several of the HLA-associated mutations lie at positions associated with drug resistance, our results indicate possible negative effects of HLA genotypes on the development of HIV-1 drug resistance.**

Human immunodeficiency virus (HIV) infection has become a major global human health issue, with more than 39 million people infected worldwide and 2.9 million AIDS-related deaths in 2006 alone (32). A major challenge to natural or vaccine-induced immune control of HIV is the ability of the virus to mutate rapidly when it comes under pressure from the host's immune system (4, 8, 15, 26). Antiviral cytotoxic T lymphocytes (CTLs) kill HIV-infected cells upon the recognition of specific viral epitopes. HIV type 1 (HIV-1) escape mutations interfere with the processing of viral antigens by proteasomes (2, 33, 34) or evolve at critical binding sites within the human leukocyte antigen (HLA)-restricted CTL epitope, thereby abrogating binding to the HLA molecule or inhibiting efficient recognition by the T-cell receptor (13, 28). Thus, HIV escapes antiviral immune responses and eradication by the host's immune system. Such selection pressure as well as viral adaptation to antiretroviral drugs should lead to consistent changes in

the amino acid sequence of the dominant population of the viral quasispecies.

Recently, Moore et al. studied the selection pressure exerted by HLA-restricted immune responses on the evolution of the HIV-1 sequence at the population level (30). A cohort of 473 HIV-1-infected patients was genotyped for the HLA-A and HLA-B loci. The most recent sequence of the HIV-1 reverse transcriptase (RT) between amino acid positions 20 and 227 was aligned to an HIV-1 consensus sequence, and viral mutations were identified. These mutations were then tested for association with distinct HLA-A or -B alleles. The authors identified 64 positive and 17 negative associations, although only 12 remained after correction for multiple testing. Several of these mutations were located in known CTL epitopes.

In a second study of the same cohort, the aforementioned group identified interactions between antiretroviral drugs and HLA alleles and diversity in the RT and protease viral sequences (21). These interactions led to higher frequencies of antiviral drug resistance mutations in patients with certain HLA alleles in some cases but also to lower frequencies in other cases. This indicates that HLA-dependent specific immune responses can support but also prevent the evolution of drug resistance.

\* Corresponding author. Mailing address: Medizinische Klinik 1, Universitätsklinik Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany. Phone: 49 228 2876789. Fax: 49 228 2879822. E-mail: ulrich.spengler@ukb.uni-bonn.de.

† Both authors contributed equally to this work.

∇ Published ahead of print on 22 August 2007.

The previous studies have analyzed the HLA-driven evolution of HIV-1 in only a fragment of the RT and protease. Therefore we wanted to examine if this phenomenon can be confirmed in the entire first half of the RT. We were also interested in extending the analysis to include the major histocompatibility complex (MHC) class II locus HLA-DRB1 to better understand selection pressure by CD4<sup>+</sup> T helper cells at the population level. In order to minimize the influence of founder effects on the HLA associations found (7), we limited the analysis to only those patients infected by HIV-1 clade B viruses and performed an analysis of potential viral lineage effects within the cohort.

Furthermore, we wanted to assess the presence of “hot spots,” where the sequence mutates more easily/rapidly due to immune pressure, and how mutations persist over time.

Finally, to understand the clinical significance of our findings, we analyzed whether HLA-driven mutations in the RT and/or protease sequence of HIV-1 lead to antiviral drug resistance and if a patient’s HLA type has an impact on whether drug resistance mutations are accumulated in a specific order in the case of thymidine analogue mutations (TAMs).

**MATERIALS AND METHODS**

**Patients.** We studied 179 clade B HIV-1-positive patients being treated at a single hospital in Bonn, Germany. The patients were monitored every 3 months, and any complications were recorded and classified according to the European modification of the 1986 Centers for Disease Control and Prevention staging (11). Antiviral therapy was advised according to updated recommendations on antiretroviral treatment for HIV infection from the relevant International AIDS Society panel. HIV-1 sequence data were collected between March 1999 and May 2003, with some patients having sequences collected at more than one time point. Additional information on HIV transmission, ethnicity, sex, and hepatitis B virus (HBV) and HCV infection status was also acquired (Table 1).

HLA genotyping was performed using the INNO-LiPA line probe assays from Innogenetics for HLA-A, HLA-B (MHC class I loci), and HLA-DRB1 (MHC class II locus) according to the manufacturer’s instructions. These are assays based on the reverse hybridization principle (12). Ambiguous results were further resolved by strand-specific PCR (One Lambda, Canoga Park, CA).

For HIV genotyping, a fragment of the *pol* gene containing the complete protease gene and the first 650 to 750 nucleotides of the RT gene was analyzed by direct sequencing of PCR products, as described in the work of Balduin et al. (3).

All patient and sequence information was entered into a relational database specifically designed for the analysis of the host immune response effects in HIV-1 infection.

**HLA-associated mutations in the RT and protease.** All amino acids in the complete protease and between positions 1 and 330 in the RT were examined in the most recent sequences from all patients. As a first step, we analyzed each amino acid position by use of an extension of Fisher’s exact test for associations with HLA alleles (29). In order to improve the power of the calculations, very rare alleles (less than or equal to 4% of the cohort) were excluded from the analysis. All HLA allele covariates with *P* values of less than 0.05 were identified and fitted in a subsequent multivariate analysis to logistic regression models. We used binomial models where the linear predictor consisted of all significant alleles and the response was a factor for which variant amino acids were classified as successes. Correction for multiple testing used the false discovery rate (FDR) method (6). Due to the low numbers of deletions and insertions in the sequences, no separate analysis was done to take these into account.

**Distribution and persistence of HLA-associated mutations.** A previous study has described CTL epitope “hot spots,” in which immunodominant epitopes cluster within distinct regions of the HIV gp120 protein (9). The CTL epitopes collected in the HIV Molecular Immunology Database at Los Alamos also appeared to be localized in particular regions of HIV proteins (27). We were interested in determining whether the distribution of HLA-associated mutations was uniform across the RT and protease. Therefore, the distribution of known mutations was compared to 10,000 randomly generated distributions according

TABLE 1. Characteristics of the patient cohort

Parameter <sup>a</sup>	Value <sup>b</sup>
Total patients (no.)	179
Age (yr)	
Average	42
Range	22–70
Sex (no. of females/no. of males)	15/164
Ethnicity	
Caucasian	161 (90.0)
Other	13 (7.3)
Unknown	5 (2.7)
Risk code	
Homosexual	99 (55.3)
Heterosexual	12 (6.7)
i.v. drug abuse	12 (6.7)
Residency in region of epidemicity	3 (1.7)
Exposure to blood products	1 (0.5)
Hemophilia	40 (22.3)
Unknown	12 (6.7)
HIV-1 viral subtype	
B	179 (100)
Chronic HCV infection	55 (30.7)
HCV genotype	
1	29 (16.2)
2	8 (4.5)
3	10 (5.6)
4	2 (1.1)
Multiple	1 (0.5)
Unknown	5 (2.7)
HBV status	
HBs+	15 (8.4)
Anti-HBs–, anti-HBc+	23 (12.8)
Anti-HBs+, Anti-HBc+	45 (25.1)
Anti-HIV therapy status	
Naive	14 (7.8)
ART (<3 antiviral drugs)	18 (10.1)
HAART (≥3 antiviral drugs)	147 (82.1)

<sup>a</sup> i.v., intravenous; ART, antiretroviral therapy.

<sup>b</sup> Unless otherwise indicated, values are numbers (percentages) of patients.

to the uniform model, each with an equal number of sequence mutations, and tested for statistical significance.

We also analyzed the persistence of HLA-associated mutations over time in our cohort. For 70 patients, HIV sequences were available for at least two individual time points that were a minimum of 6 months apart (the sequences for 50 patients were collected more than 1 year apart). None of these sequences was acquired during the acute stage of HIV infection; rather, they were acquired at a later time point. The persistence of mutations was examined at each of the already identified positions.

**Impact of HLA on drug resistance mutation pathways.** During antiretroviral therapy, the virus is exposed to strong selective pressure, which can result in an accumulation of mutations conferring drug resistance. These mutations are usually persistent, provided that there is continuous drug-induced selection pressure (31). Therefore, the development of resistance within the HIV genome can be regarded as the accumulation of such mutations. This accumulation has been modeled by weighted branchings or directed trees, which provide an intuitive model of directed dependencies between events and their time of occurrence. The single-tree model has been extended to mixtures of trees (so-called mutagenetic tree mixture models) in order to capture more complex evolutionary scenarios, for which the software package Mtreemix has been developed (5).

Of particular interest were the TAMs in the RT that can arise after treatment with nucleoside RT inhibitors (NRTIs) zidovudine, stavudine, and abacavir. Studies have suggested that HIV-1 develops TAMs by one of two distinct pathways: TAM1 (41L, 210W, and 215F/Y) or TAM2 (67N, 70R, and 219E/Q) (17).

We identified all patients in our cohort that had undergone NRTI treatment and examined their HIV sequences for TAMs to determine whether particular HLA alleles can be associated with either the TAM1 or the TAM2 pathway.

**HLA-driven selection at antiretroviral drug resistance sites.** Several HLA allele-specific mutations have been reported to be located in positions of known drug resistance mutations (21, 22). An analysis of the sequences of our patient

cohort was performed in order to determine whether similar associations could be identified.

**Phylogenetic analysis.** In order to explore the impact of viral lineage founder effects, we applied ProtTest version 1.3 (1, 14, 16) to find the best-fitting model of protein evolution for the HIV sequence alignment of the cohort. The best-fitting model according to both the Akaike information criterion and the Bayesian information criterion (19) was the Jones-Taylor-Thornton (JTT) model (23), with gamma rates, variable amino acid frequencies, and invariable sites. We then used PhyML (16) for estimating a maximum-likelihood phylogeny for the given protein evolution model and performed 100 bootstrap replications in order to obtain bootstrap support values. Subsequently, we used neighbor-net (10, 20) to get a better visualization of the noise in the phylogenetic signal.

## RESULTS

**Patients.** During the observation period, 147 (82.1%) patients were treated with highly active antiretroviral therapy (HAART), which was based on a protease inhibitor (PI; indinavir, saquinavir, nelfinavir, amprenavir, or ritonavir) for 117 patients and on a nonnucleoside RT inhibitor (NNRTI; nevirapine, efavirenz, or delavirdine) for 30 patients. HAART was the first-line treatment for 61 patients, and 86 had previously undergone therapy with nucleoside analogues. HAART was changed to different drug combinations for 85 of 117 patients who received HAART based on PIs and for 23 of 30 patients who received HAART based on NNRTIs to address side effects or emerging viral resistance.

The antiretroviral treatment group consisted of 18 (10.1%) patients who received two or fewer antiretroviral drugs. One patient received zidovudine monotherapy, and 17 were treated with several nucleoside analogues (zidovudine, didanosine, zalcitabine, stavudine, or lamivudine). The remaining 14 patients (7.8%) did not receive any antiretroviral drugs.

With regard to the length of treatment, 18 patients had been treated for <1 year, 62 patients had been treated for between 1 and 5 years, and 88 patients had been treated for >5 years. The staging of HIV disease according to the European modification of the 1986 Centers for Disease Control and Prevention staging (11) in the cohort was A for 58 patients, B for 59 patients, and C for 57 patients. Five patients were not classified.

**HLA-associated mutations in the RT and protease.** HIV RT sequences (amino acid positions 1 to 330) were initially aligned with the reference sequence HXB2 (27). Associations were subsequently confirmed using the population consensus sequence, which was generated by assigning the most common amino acid for each position of all sequences pooled from the cohort. Overall, 15 associations with uncorrected *P* values of less than 0.005 were found; these also had *P* values of less than 0.05 in the logistic regression models (Table 2). Ten of these associations had *P* values of less than 0.05 after correction for multiple testing.

We were able to confirm four of the associations that Moore et al. reported as significant after correction for the total number of residues examined across the entire region (30): HLA-B\*51 at amino acid position 135, B\*07 at position 162, A\*11 at position 166, and B\*35 at position 177. We also found a number of novel HLA allele-specific associations, with four of those associations lying outside the part of the RT sequence analyzed by Moore et al. Seven of the mutations in the RT sequence are located in previously defined epitopes, with six of these being located in known epitope anchor positions.

Interestingly, six of the identified associations were negative associations, indicating that mutations in the RT were less likely if the patient carried that particular allele: amino acid position 177 was negatively associated with HLA-B\*35, position 178 with HLA-B\*35, position 188 with HLA-DRB1\*12, position 207 with B\*15, position 277 with HLA-A\*03, and position 291 with HLA-B\*27.

In the protease, we found HLA-associated mutations at seven positions with uncorrected *P* values of less than 0.005; these also had *P* values of less than 0.05 in the logistic regression models (Table 3). Four associations had *P* values of less than 0.05 after correction for multiple testing, with three of these associations being located in previously defined epitopes. No negative associations were found.

For the two HLA-DRB1-associated mutations in the RT (position 67 for HLA-DRB1\*08 and position 188 for HLA-DRB1\*12), the alleles are not known to be in linkage disequilibrium with any class I alleles which were included in the multivariate analysis at each position.

Overall, HLA-B alleles were involved in more associations ( $n = 15$  [68%]) than alleles from either HLA-A ( $n = 5$  [23%]) or HLA-DRB1 ( $n = 2$  [9%]).

**Distribution and persistence of HLA-associated mutations.** The HLA-associated mutations that have been identified in our patient cohort do not appear to be uniformly distributed across the protease and RT but are more frequent in regions known to have many epitopes. However, an analysis of the nonrandom distribution of the HIV-1 sequence mutations showed only a weak level of significance in the protease (*P* value equal to 0.06) and none in the RT (data not shown).

For most patients, sequence mutations were consistently found in all available sequences. In relatively few cases, the mutations were not persistent, with the wild-type amino acid being replaced by a variant amino acid in most cases (Tables 2 and 3). Drug therapy did not appear to influence the gain or loss of HIV-1 sequence mutations, as most patients had HAART composed of diverse drug combinations and yet their mutations remained stable throughout.

**Impact of HLA on drug resistance mutation pathways.** Of our patient cohort, 165 patients were treated with an NRTI regimen containing zidovudine, stavudine, or abacavir before the sequence used for analysis was generated. Of these patients, 72 had no TAMs reported. For TAM1 (41L, 210W, and 215F/Y), 9 patients matched the pathway and 25 showed either one mutation too few or one too many. For TAM2 (67N, 70R, and 219E/Q), 10 patients had matching mutations and 10 patients showed either one mutation too many or one too few. The remaining patients showed even rarer combinations of these mutations.

We looked for associations between either the TAM1 or the TAM2 pathway and a particular HLA type by use of Mtreemix. This also allowed for imperfect matches to a particular pathway, because the software supports the analysis of complex evolutionary scenarios. However, no significant associations were found.

**HLA-driven selection at antiretroviral drug resistance sites.** We examined all associations for our cohort, which are also known drug resistance mutations (Table 4). Our results differ somewhat from those of John et al. (21), who reported drug resistance mutations at protease amino acid residues 20, 32, 36,

TABLE 2. HIV-1 sequence mutations in the HIV-1 RT

Polymorphism <sup>a</sup>	HLA type <sup>b</sup>	No. of patients with HIV-1 sequence mutation:		OR <sup>d</sup>	P value		Known epitopes <sup>g</sup>	Persistence <sup>h</sup>	New <sup>i</sup>
		Absent <sup>c</sup>	Present		Unadjusted <sup>e</sup>	Adjusted <sup>f</sup>			
E6x	HLA-B*40 Non-HLA-B*40	23 139	11 6	11.08	0.00001	0.00047	B*4001; aa 5-12 (ETV <sup>h</sup> VVKL)	Persistent, 9; multiple pops, 3; definite change, 0	Y
V35x	HLA-A*24 Non-HLA-A*24	19 112	17 31	3.23	0.00314	>0.05		Persistent, 6; multiple pops, 1; definite change, 1	Y
D67x	HLA-DRB1*08 Non-HLA-DRB1*08	1 120	6 52	13.85	0.00501	>0.05		Persistent, 3; multiple pops, 2; definite change, 0	Y
V75x	HLA-B*57 Non-HLA-B*57	6 154	5 14	9.17	0.00244	>0.05		Persistent, 4; multiple pops, 2; definite change, 1	Y
I135x	HLA-B*51 Non-HLA-B*51	2 97	22 58	18.40	<0.00001	0.00002	B*5101/B51; aa 128-135 (TAF <sup>h</sup> TTPSI)	Persistent, 8; multiple pops, 0; definite change, 1	N
E138x	HLA-A*24 Non-HLA-A*24	31 143	5 0		0.00026	0.00859		Persistent, 7; multiple pops, 0; definite change, 1	Y
S162x	HLA-B*07 Non-HLA-B*07	21 108	27 23	6.04	<0.00001	0.00004	B7; aa 153-165 (WKG <sup>h</sup> PAIFQSSMT), aa 153-165 (WKG <sup>h</sup> SPAI <sup>h</sup> FQSSMT), aa 156-164 (SPAI <sup>h</sup> FQSSMT), aa 156-165 (SPAI <sup>h</sup> FQSSMT)	Persistent, 14; multiple pops, 1; definite change, 3	N
K166x	HLA-A*11 Non-HLA-A*11	13 152	7 7	11.69	0.00016	0.00521	A*1101/A11; aa 158-166 (AIF <sup>h</sup> OSSM TK), aa 158-166 (SHF <sup>h</sup> OSSM TK)	Persistent, 5; multiple pops, 0; definite change, 0	N
D177x	HLA-B*35 Non-HLA-B*35	16 125	18 20	0.14	<0.00001	0.00013	B*3501/B35; aa 175-183 (NPD <sup>h</sup> IVY OY); aa 175-183 (HPD <sup>h</sup> IVY OY)	Persistent, 10; multiple pops, 0; definite change, 1	N
I178x	HLA-B*35 Non-HLA-B*35	22 129	12 16	0.23	0.00125	0.04121	B*3501/B35; aa 175-183 (NPD <sup>h</sup> IVY OY); aa 175-183 (HPD <sup>h</sup> IVY OY)	Persistent, 10; multiple pops, 0; definite change, 1	Y
Y188x	HLA-DRB1*12 Non-HLA-DRB1*12	5 167	3 4	0.04	0.00191	>0.05		Persistent, 3; multiple pops, 0; definite change, 1	Y
Q207x	HLA-B*15 Non-HLA-B*15	10 111	20 38	0.17	0.00003	0.00090		Persistent, 8; multiple pops, 1; definite change, 2	Y
K277x	HLA-A*03 Non-HLA-A*03	1 89	38 51	0.02	<<0.00001	<<0.00001	A3; aa 269-277 (QI <sup>h</sup> YPIG <sup>h</sup> KV <sup>h</sup> R)	Persistent, 14; multiple pops, 1; definite change, 1	Y
K277x	HLA-B*44 Non-HLA-B*44	26 64	9 80	3.61	0.00225	0.02616		Persistent, 12; multiple pops, 2; definite change, 1	Y
E291x	HLA-B*27 Non-HLA-B*27	10 163	3 3	0.06	0.00534	>0.05		Persistent, 4; multiple pops, 0; definite change, 0	Y

<sup>a</sup> The variant position relative to the wild-type amino acid.<sup>b</sup> The specific HLA type for which a mutation was found.<sup>c</sup> Patients without the indicated HIV-1 sequence mutation had the corresponding wild-type amino acid.<sup>d</sup> OR, odds ratio.<sup>e</sup> The unadjusted *P* value was from the contingency table and was determined by use of Fisher's exact test.<sup>f</sup> The adjusted *P* value was determined using FDR.<sup>g</sup> Any known epitopes for the HLA alleles and their sequences (27), aa, amino acid.<sup>h</sup> The persistence of the HIV-1 sequence mutation in individual patients at a minimum of two different time points. Persistent, all dominant viral sequences from a patient were identical at the different time points; multiple pops, dominant viral sequences identified in a patient were different at the various time points without a consistent change from one type to another; definite change, the patient's dominant viral sequence showed a clear switch at this position from one amino acid to another amino acid. Persistent, multiple pops, and definite change numbers are numbers of patients.<sup>i</sup> Y indicates that the HIV-1 sequence mutation is new; N indicates that it has already been described in the literature.

TABLE 3. HIV-1 sequence mutations in the HIV-1 protease

Polymorphism <sup>a</sup>	HLA type <sup>b</sup>	No. of patients with HIV-1 sequence mutation:		OR <sup>d</sup>	P value		Known epitopes <sup>g</sup>	Persistence <sup>h</sup>	New <sup>i</sup>
		Absent <sup>c</sup>	Present		Unadjusted <sup>e</sup>	Adjusted <sup>f</sup>			
		E35x	HLA-B*44		11	24			
	Non-HLA-B*44	120	24						
E35x	HLA-A*68	11	13	4.05	0.00238	0.03922	A*6802; aa 30–38 (DTVLEDI NL), aa 30–38 (DTVLEEM NL), aa 30–38 (DTVLEE WNL)	Persistent, 7; multiple pops, 0; definite change, 0	Y
	Non-HLA-A*68	120	35						
N37x	HLA-B*44	9	26	6.15	<0.00001	0.00022	B*44; aa 34–42 (EEMNLPGRW)	Persistent, 11; multiple pops, 3; definite change, 1	Y
	Non-HLA-B*44	98	46						
I54x	HLA-B*57	5	6	7.57	0.00285	>0.05		Persistent, 4; multiple pops, 0; definite change, 3	Y
	Non-HLA-B*57	145	23						
V82x	HLA-B*57	5	6	6.55	0.00499	>0.05		Persistent, 5; multiple pops, 1; definite change, 1	Y
	Non-HLA-B*57	142	26						
Q92x	HLA-B*15	22	8	17.70	0.00003	0.00099		Persistent, 10; multiple pops, 1; definite change, 0	Y
	Non-HLA-B*15	146	3						
I93x	HLA-B*15	11	19	3.63	0.00184	>0.05		Persistent, 10; multiple pops, 0; definite change, 1	Y
	Non-HLA-B*15	101	48						

<sup>a</sup> The variant position relative to the wild-type amino acid.

<sup>b</sup> The specific HLA type for which a mutation was found.

<sup>c</sup> Patients without the indicated HIV-1 sequence mutation had the corresponding wild-type amino acid.

<sup>d</sup> OR, odds ratio.

<sup>e</sup> The unadjusted *P* value was from the contingency table and was determined by use of Fisher's exact test.

<sup>f</sup> The adjusted *P* value was determined using FDR.

<sup>g</sup> Any known epitopes for the HLA alleles and their sequences (27), aa, amino acid.

<sup>h</sup> The persistence of the HIV-1 sequence mutation in individual patients at a minimum of two different time points. Persistent, all dominant viral sequences from a patient were identical at the different time points; multiple pops, dominant viral sequences identified in a patient were different at the various time points without a consistent change from one type to another; definite change, the patient's dominant viral sequence showed a clear switch at this position from one amino acid to another amino acid. Persistent, multiple pops, and definite change numbers are numbers of patients.

<sup>i</sup> Y indicates that the HIV-1 sequence mutation is new; N indicates that it has already been described in the literature.

and 48 to have positive HLA associations. For our cohort, we could not confirm any of these associations. John et al. also reported the RT amino acid residues 41, 67, 70, 118, 210, and 215 to have positive HLA associations matching NRTI-associ-

ated mutations. Of these, only residue 67 was associated with DRB1\*08 in our cohort (A\*10 in the work of John et al. [21]).

We also found eight new associations, five in the protease and three in the RT. For three of the associations in the

TABLE 4. HIV-1 sequence mutations at positions which are also known drug resistance mutations and numbers of patients exhibiting the indicated HLA type who received the indicated drug type

Drug or resistance type	Polymorphism <sup>a</sup>	HLA type <sup>b</sup>	No. of patients with HIV-1 sequence mutation:		No. of patients (%) who received drugs	New <sup>d</sup>
			Absent <sup>c</sup>	Present		
			Tipranavir-ritonavir	PRO-E35x		
		Non-HLA-B*44	120	24	0 of 155 (0)	
	PRO-E35x	HLA-A*68	11	13	0 of 35 (0)	Y
		Non-HLA-A*68	120	35	0 of 144 (0)	
Many currently used PIs	PRO-I54x	HLA-B*57	5	6	10 of 11 (90.9)	Y
		Non-HLA-B*57	145	23	117 of 168 (69.6)	
	PRO-V82x	HLA-B*57	5	6	10 of 11 (90.9)	Y
		Non-HLA-B*57	142	26	117 of 168 (69.6)	
Atazanavir	PRO-Q92x	HLA-B*15	22	8	0 of 30 (0)	Y
		Non-HLA-B*15	146	3	0 of 149 (0)	
Multi-NRTI resistance	RT-D67x	HLA-DRB1*08	1	6	7 of 7 (100)	Y
		Non-HLA-DRB1*08	120	52	158 of 172 (91.8)	
	RT-V75x	HLA-B*57	6	5	11 of 11 (100)	Y
		Non-HLA-B*57	154	14	154 of 168 (91.6)	
NNRTIs	RT-Y188x	HLA-DRB1*12	5	3	5 of 8 (62.5)	Y
		Non-HLA-DRB1*12	167	4	84 of 171 (49.1)	

<sup>a</sup> HIV-1 sequence mutation in the protease or RT.

<sup>b</sup> Specific HLA type for which the mutation was found.

<sup>c</sup> Patients without the indicated HIV-1 sequence mutation had the corresponding wild-type amino acid.

<sup>d</sup> Y indicates that the association is new; N indicates that it has already been described in the literature.

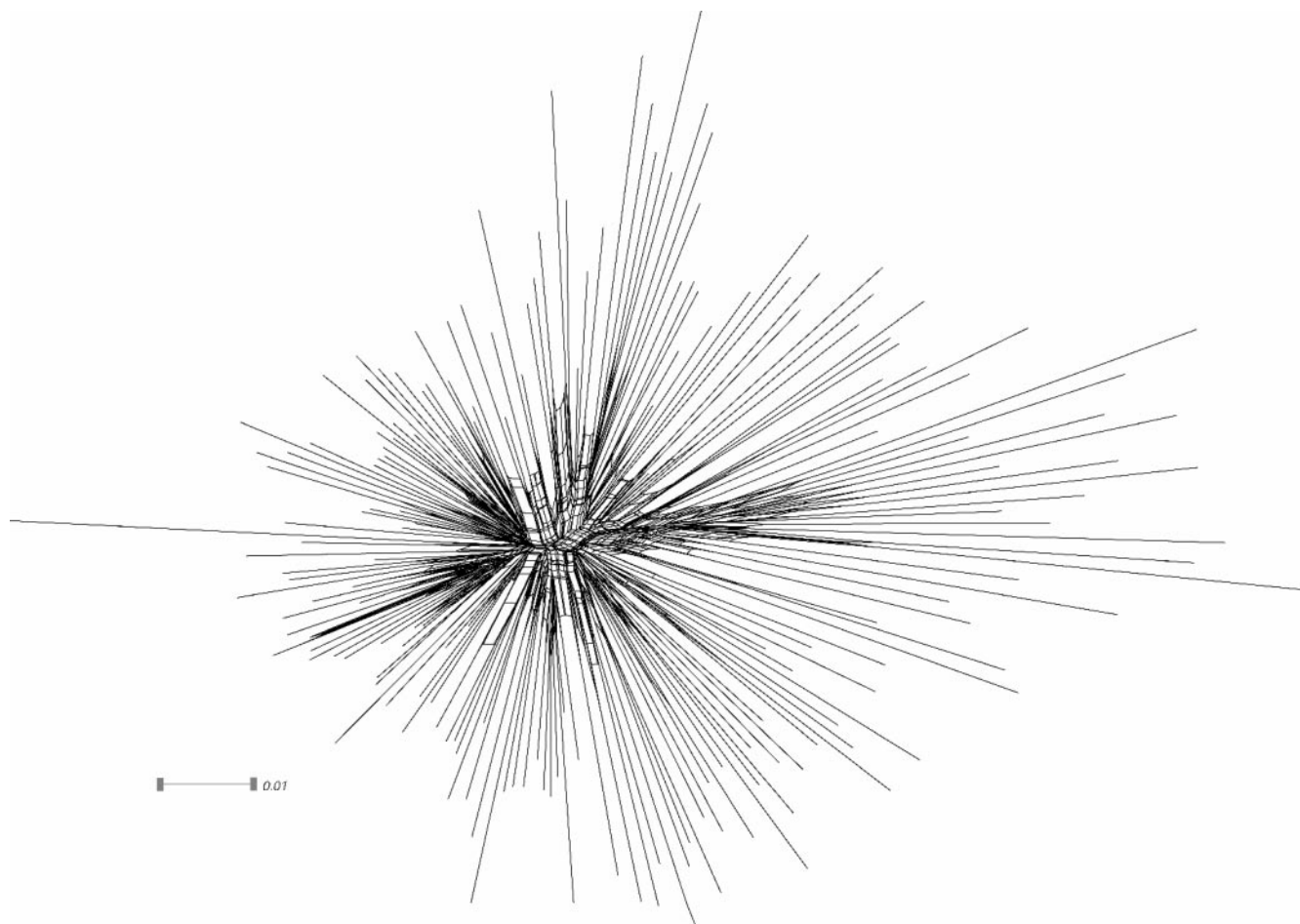


FIG. 1. Neighbor-net network of the cohort illustrating a star-like shape and a high level of netting in the center of the phylogeny.

protease (residue 35 for A\*68 and B\*44 and residue 92 for B\*15), none of the patients in our cohort had ever been treated with the relevant antiviral therapy (tipranavir-ritonavir or atazanavir). In contrast, for the remaining five associations (residues 54 and 82 in the protease and residues 67, 75, and 188 in the RT), most patients had been treated with the relevant antiviral therapy. These residues are associated with resistance against most currently available PIs, NNRTIs, and multinucleoside and -nucleotide RT inhibitors (multi-NRTIs). However, statistically significant differences for patients taking these drugs could not be found between individuals with or without the HLA allele.

We observed that HCV-coinfected patients were less likely to have received antiretroviral therapy which included PIs.

**Phylogenetic analysis.** The shape of the maximum-likelihood phylogeny is almost star-like: the sequence diversity within clusters is higher than the diversity between clusters. Another indicator for the star likeness is the low bootstrap support for all branches in the interior of the phylogeny. The resulting neighbor-net network (Fig. 1) shows a good deal of netting in the center of the phylogeny and rejects a clustering into distinct groups. Hence, the neighbor-net method also suggests a star-like shape for the phylogeny. Given that the sequences evolved along a star-like phylogeny, we can rule out

founder effects or other artifacts due to a shared evolutionary history.

## DISCUSSION

Selection of the dominant HIV-1 population among the viral quaspecies is likely to reflect the combined effects of viral adaptation to the host's immune response and antiretroviral drugs. MHC class I alleles have been described as affecting the evolution of the HIV-1 sequence on the population level in a single previous cohort from Western Australia (21, 30). We were interested in examining whether these findings could be confirmed for our Western European cohort.

Four out of a total of 12 significant associations previously reported for the RT were confirmed for our cohort, all of which corresponded to known epitopes (Table 2). Moreover, we describe 11 new associations, with 4 of these lying outside the region of the RT analyzed previously (30). Several of these associations also correspond to known T-cell epitopes.

The identified CTL escape mutations, for which known epitopes exist, do not consistently lie in known anchor positions. Due to the complex pathway by which an epitope is processed before being displayed on the cell surface by the HLA molecule and the incompletely understood nature of

T-cell receptor binding and T-cell activation, HIV-1 sequence mutations in nonanchor positions could very well still have a negative impact on the recognition of such epitopes and hence be beneficial for the virus.

These data support the hypothesis that the associations found are indeed epitopes targeted by CTLs and that continuous pressure by CTL responses leads to the selection of survival strains that have mutations in these regions. The fact that we cannot reproduce all associations described in the previous publications and that we found a number of additional associations may be due to several reasons: the first cohort consisted of patients from Western Australia, whereas our patients are mainly Caucasians from Germany. Thus, the available HLAs may differ, and the levels of access to and types of antiviral drugs may be different. However, the fact that we were able to confirm a number of associations despite these differences proves that our main conclusions, although affected by local factors, are generally true. Furthermore, HLA-associated mutations, once acquired, were stable and found in later viral isolates from the same patient. This supports the hypothesis that the continuing selection pressure by T-cell recognition prevents the virus from reverting back despite changes in the antiviral drug regimen.

MHC class I-dependent CTL responses shape the viral sequence by directly killing infected cells that present the right epitope in their HLA molecules. So far, the impact of MHC class alleles II on viral evolution has been studied only for a small patient group. Harcourt et al. (18) described HIV sequence variation in the p24 GAG epitope of HLA-DR1. Our data from a large patient cohort support their observation of a role of MHC class II alleles and CD4 T-cell responses in the evolution of the HIV-1 sequence. We found two associations in the RT region, also demonstrating that even class II alleles can exert selection pressure on viral sequences.

We also studied the effect of MHC class I and class II on the evolution of the full-length protease sequence. Interestingly, we did not find any of the associations which were described for the Western Australian cohort (21). The underlying reason for this is unclear. However, since most patients in our cohort and the Western Australian cohort were undergoing drug treatment, it is possible that the type, time point, and length of use of certain PIs may have affected the results from our study as well those from the previous studies. Nevertheless, three associations were found within previously described epitopes for the same HLA allele, and four associations remained significant after correction for multiple testing, supporting the validity of the associations reported.

Although we were not able to confirm statistically the existence of “hot spots,” as the effect reached only a weak level of statistical significance ( $P$  value equal to 0.06), it appears that certain regions in the RT and protease are targeted by a greater number of epitopes and therefore CTLs, thus driving viral evolution in this region.

Recently, a study by Kiepiela et al. reported that the relative contribution of HLA-B alleles outweighs the contribution of HLA-A alleles in influencing HIV disease outcome (25). This is also reflected in our results, as more significant associations were found with HLA-B alleles ( $n = 15$  [68%]) than with HLA-A alleles ( $n = 5$  [23%]) or HLA-DRB1 alleles ( $n = 2$  [9%]).

For most patients, HLA-associated HIV-1 sequence mutations were consistently found for all sequences from that patient, while only relatively few cases had HLA-associated mutations that were not stable. These changes were generally from a wild-type amino acid to the variant amino acid in the subsequent sample. This observation is consistent with other studies showing that CTL escape mutations develop soon after infection (24).

We also analyzed whether these HLA-associated mutations were influenced by drug therapy and found that the impact of drug therapy on the associations for our patient cohort was not statistically significant. In some cases, patients in the cohort had not received a particular drug treatment (tipranavir or atazanavir) yet showed HLA-associated mutations at positions where known mutations providing resistance to these particular drugs can also occur. As tipranavir and atazanavir were first approved for antiviral therapy shortly after our patients' viral sequences were acquired, it is not possible for these patients to have been infected with an already resistant viral strain. This indicates either that these mutations have become locked in our population, so that the virus cannot revert back, or that HLA-driven selection of mutations may predispose some patients to the development of some drug resistance mutations. In order to optimize the therapy of patients carrying particular HLA types, the choice of future therapy regimens should take into account the fact that some patients are more likely to develop particular HLA-associated mutations, which coexist at residues of drug resistance mutations. This consideration could be made for positions in the protease that are linked to specific drugs (tipranavir and atazanavir) but not to the majority of PIs, multi-NRTIs, or NNRTIs.

In summary, we were able to confirm immune-driven selection pressure not only by MHC class I alleles but also by MHC class II alleles on the development of escape mutations at a population level. Further, we have described a number of HLA-associated mutations in the RT as well as protease and, finally, have analyzed their possible impact on the success of antiviral drug treatment.

#### ACKNOWLEDGMENTS

We thank Monika Schulz for help with the HLA typing and building the patient database and Jochen Maydt, Jörg Rahnenführer, and Tobias Sing for constructive suggestions for this work.

#### REFERENCES

1. Abascal, F., R. Zardoya, and D. Posada. 2005. ProfTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**:2104–2105.
2. Allen, T. M., M. Altfeld, X. G. Yu, K. M. O'Sullivan, M. Lichterfeld, S. Le Gall, M. John, B. R. Mothe, P. K. Lee, E. T. Kalife, D. E. Cohen, K. A. Freedberg, D. A. Strick, M. N. Johnston, A. Sette, E. S. Rosenberg, S. A. Mallal, P. J. Goulder, C. Brander, and B. D. Walker. 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.
3. Balduin, M., S. Sierra, M. P. Daumer, J. K. Rockstroh, M. Oette, G. Fatkenheuer, B. Kupfer, N. Beerenwinkel, D. Hoffmann, J. Selbig, H. J. Pfister, and R. Kaiser. 2005. Evolution of HIV resistance during treatment interruption in experienced patients and after restarting a new therapy. *J. Clin. Virol.* **34**:277–287.
4. Barouch, D. H., J. Kunstman, M. J. Kuroda, J. E. Schmitz, S. Santra, F. W. Peyerl, G. R. Krivulka, K. Beaudry, M. A. Lifton, D. A. Gorgone, D. C. Montefiori, M. G. Lewis, S. M. Wolinsky, and N. L. Letvin. 2002. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**:335–339.
5. Beerenwinkel, N., J. Rahnenführer, R. Kaiser, D. Hoffmann, J. Selbig, and T. Lengauer. 2005. Mtree: a software package for learning and using mixture models of mutagenetic trees. *Bioinformatics* **21**:2106–2107.

6. **Benjamini, Y., and Y. Hochberg.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**: 289–300.
7. **Bhattacharya, T., M. Daniels, D. Heckerman, B. Foley, N. Frahm, C. Kadie, J. Carlson, K. Yusim, B. McMahon, B. Gaschen, S. Mallal, J. I. Mullins, D. C. Nickle, J. Herbeck, C. Rousseau, G. H. Learn, T. Miura, C. Brander, B. Walker, and B. Korber.** 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* **315**:1583–1586.
8. **Borrow, P., H. Lewicki, X. Wei, M. S. Horwitz, N. Peffer, H. Meyers, J. A. Nelson, J. E. Gairin, B. H. Hahn, M. B. Oldstone, and G. M. Shaw.** 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.
9. **Brown, S. A., T. D. Lockey, C. Slaughter, K. S. Slobod, S. Surman, A. Zirkel, A. Mishra, V. R. Pagala, C. Coleclough, P. C. Doherty, and J. L. Hurwitz.** 2005. T cell epitope “hotspots” on the HIV type 1 gp120 envelope protein overlap with tryptic fragments displayed by mass spectrometry. *AIDS Res. Hum. Retrovir.* **21**:165–170.
10. **Bryant, D., and V. Moulton.** 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**:255–265.
11. **Centers for Disease Control and Prevention.** 1986. Classification system for human T-lymphotropic virus type III/lymphadenopathy-associated virus infections. *Morb. Mortal. Wkly. Rep.* **35**:334–339.
12. **De Vreese, K., R. Barylski, F. Pughe, M. Blaser, C. Evans, J. Norton, G. Semana, R. Holman, P. Loiseau, D. Masson, M. Gielis, A. De Brauwier, I. De Canck, G. Verpooten, and F. Hulstaert.** 2004. Performance characteristics of updated INNO-LiPA assays for molecular typing of human leukocyte antigen A (HLA-A), HLA-B, and HLA-DQB1 alleles. *Clin. Diagn. Lab. Immunol.* **11**:430–432.
13. **Draenert, R., S. Le Gall, K. J. Pfafferoth, A. J. Leslie, P. Chetty, C. Brander, E. C. Holmes, S. C. Chang, M. E. Feeney, M. M. Addo, L. Ruiz, D. Ramduth, P. Jeena, M. Altfeld, S. Thomas, Y. Tang, C. L. Verrill, C. Dixon, J. G. Prado, P. Kiepiela, J. Martinez-Picado, B. D. Walker, and P. J. Goulder.** 2004. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J. Exp. Med.* **199**:905–915.
14. **Drummond, A., and K. Strimmer.** 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17**: 662–663.
15. **Goulder, P. J., R. E. Phillips, R. A. Colbert, S. McAdam, G. Ogg, M. A. Nowak, P. Giangrande, G. Luzzi, B. Morgan, A. Edwards, A. J. McMichael, and S. Rowland-Jones.** 1997. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**:212–217.
16. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
17. **Hanna, G. J., V. A. Johnson, D. R. Kuritzkes, D. D. Richman, A. J. Brown, A. V. Savara, J. D. Hazelwood, and R. T. D’Aquila.** 2000. Patterns of resistance mutations selected by treatment of human immunodeficiency virus type 1 infection with zidovudine, didanosine, and nevirapine. *J. Infect. Dis.* **181**:904–911.
18. **Harcourt, G. C., S. Garrard, M. P. Davenport, A. Edwards, and R. E. Phillips.** 1998. HIV-1 variation diminishes CD4 T lymphocyte recognition. *J. Exp. Med.* **188**:1785–1793.
19. **Hastie, T., R. Tibshirani, and J. Friedman.** 2001. The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, Heidelberg, Germany.
20. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**:254–267.
21. **John, M., C. B. Moore, I. R. James, and S. A. Mallal.** 2005. Interactive selective pressures of HLA-restricted immune responses and antiretroviral drugs on HIV-1. *Antivir. Ther.* **10**:551–555.
22. **Johnson, V. A., F. Brun-Vezinet, B. Clotet, B. Conway, D. R. Kuritzkes, D. Pillay, J. M. Schapiro, A. Telenti, and D. D. Richman.** 2005. Update of the drug resistance mutations in HIV-1: fall 2005. *Top. HIV Med.* **13**:125–131.
23. **Jones, D. T., W. R. Taylor, and J. M. Thornton.** 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
24. **Jones, N. A., X. Wei, D. R. Flower, M. Wong, F. Michor, M. S. Saag, B. H. Hahn, M. A. Nowak, G. M. Shaw, and P. Borrow.** 2004. Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response. *J. Exp. Med.* **200**:1243–1256.
25. **Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferoth, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. Goulder.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**:769–775.
26. **Koenig, S., A. J. Conley, Y. A. Brewah, G. M. Jones, S. Leath, L. J. Boots, V. Davey, G. Pantaleo, J. F. Demarest, C. Carter, et al.** 1995. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nat. Med.* **1**:330–336.
27. **Korber, B. T. M., C. Brander, B. F. Haynes, R. Koup, J. P. Moore, B. D. Walker, and D. I. Watkins (ed.).** 2005. HIV molecular immunology 2005. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, NM.
28. **McMichael, A. J., and S. L. Rowland-Jones.** 2001. Cellular immune responses to HIV. *Nature* **410**:980–987.
29. **Mehta, C., and N. Patel.** 1986. FEXACT: a Fortran subroutine for Fisher’s exact test on unordered r\*c contingency tables. *ACM Trans. Math. Software* **12**:154–161.
30. **Moore, C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal.** 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**:1439–1443.
31. **Shafer, R. W., R. Kantor, and M. J. J. Gonzales.** 2000. The genetic basis of HIV-1 resistance to reverse transcriptase and protease inhibitors. *AIDS Rev.* **2**:211–228.
32. **UNAIDS/WHO.** 2006. AIDS epidemic update: December 2006. *NLM classification*: WC 503.41.
33. **Yokomaku, Y., H. Miura, H. Tomiyama, A. Kawana-Tachikawa, M. Takiguchi, A. Kojima, Y. Nagai, A. Iwamoto, Z. Matsuda, and K. Ariyoshi.** 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.
34. **Zimbwa, P., A. Milicic, J. Frater, T. J. Scriba, A. Willis, P. J. Goulder, T. Pillay, H. Gunthard, J. N. Weber, H. T. Zhang, and R. E. Phillips.** 2007. Precise identification of a human immunodeficiency virus type 1 antigen processing mutant. *J. Virol.* **81**:2031–2038.