JOURNAL OF VIROLOGY, Feb. 2008, p. 1980–1992 Vol. 82, No. 4
0022-538X/08/$08.00+0 doi:10.1128/JVI.02742-06

An Integrative Bioinformatic Approach for Studying Escape Mutations in Human Immunodeficiency Virus Type 1 *gag* in the Pumwani Sex Worker Cohort▽

Harold O. Peters,[1]† Mark G. Mendoza,[1]† Rupert E. Capina,[1]† Ma Luo,[1]* Xiaojuan Mao,[1] Michael Gubbins,[2] Nico J. D. Nagelkerke,[4] Ian MacArthur,[2] Brent B. Sheardown,[2] Joshua Kimani,[3] Charles Wachihi,[3] Subo Thavaneswaran,[1] and Francis A. Plummer[1,2]

Department of Medical Microbiology, University of Manitoba, 730 William Avenue, Winnipeg, Manitoba, Canada[1]*; Public Health Agency of Canada, Winnipeg, Manitoba, Canada*[2]*; Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya*[3]*; and Department of Community Medicine, UAE University, P.O. Box 17666, Al Ain, United Arab Emirates*[4]

Received 13 December 2006/Accepted 2 November 2007

Human immunodeficiency virus type 1 (HIV-1) is able to evade the host cytotoxic T-lymphocyte (CTL) response through a variety of escape avenues. Epitopes that are presented to CTLs are first processed in the presenting cell in several steps, including proteasomal cleavage, transport to the endoplasmic reticulum, binding by the HLA molecule, and finally presentation to the T-cell receptor. An understanding of the potential of the virus to escape CTL responses can aid in designing an effective vaccine. To investigate such a potential, we analyzed HIV-1 *gag* from 468 HIV-1-positive Kenyan women by using several bioinformatic approaches that allowed the identification of positively selected amino acids in the HIV-1 *gag* region and study of the effects that these mutations could have on the various stages of antigen processing. Correlations between positively selected residues and mean CD4 counts also allowed study of the effect of mutation on HIV disease progression. A number of mutations that could create or destroy proteasomal cleavage sites or reduce binding affinity of the transport antigen processing protein, effectively hindering epitope presentation, were identified. Many mutations correlated with the presence of specific HLA alleles and with lower or higher CD4 counts. For instance, the mutation V190I in subtype A1-infected individuals is associated with HLA-B*5802 ($P = 4.73 \times 10^{-4}$), a rapid-progression allele according to other studies, and also to a decreased mean CD4 count ($P = 0.019$). Thus, V190I is a possible HLA escape mutant. This method classifies many positively selected mutations across the entire *gag* region according to their potential for immune escape and their effect on disease progression.

Cytotoxic T lymphocytes (CTLs) play a crucial role in controlling viral replication during acute and chronic human immunodeficiency virus type 1 (HIV-1) infections (15, 29, 41, 60). It is believed that during acute HIV-1 infection the initial viremia is controlled by HIV-1-specific CTLs, and these cells may suppress viral replication throughout the chronic phase of infection (11). Therefore, CTLs likely play an important role in viral control. This anti-HIV-1 immunity is influenced by the diversity of human leukocyte antigens (HLAs) (59). Infected cells present viral peptides via HLA class I molecules to HIV-1-specific CTLs, which induces effector immune responses that destroy infected cells. The response, however, is dependent on the specificity of the different HLA class I molecules. Studies have also shown that the effects of CTL responses are not equal. The CTL responses of individuals with HLA-B*5701 are associated with slower disease progression (53), while CTL responses of people with HLA-B*5802 are associated with rapid progression (44).

Proteasomal cleavage and transport antigen processing (TAP protein) transport also play crucial roles in the cellular immune response (51). Prior to peptide presentation, endogenous viral proteins undergo proteolytic cleavage into small peptides in the proteasome. There is increasing evidence that CTL-restricted epitopes are a result of C-terminal proteasomal cleavage (8, 63, 73, 82). In some epitopes, the N terminus is trimmed within the endoplasmic reticulum (ER) after the peptide has been transported by TAP, generating an 8- to 11-amino-acid fragment (8, 77). Specific peptides then bind to their corresponding HLA class I molecule and are presented on the cell surface.

Through mutations, HIV-1 can rapidly adapt to the selective pressure exerted by the host immune system. Host selective pressure accumulates viral mutations that result in escape variants. Mutational escape from recognition of CTLs may carry a fitness cost (56) or be beneficial to the virus. Therefore, the genotype and fitness of the virus is highly dependent upon the host it is infecting. This study employed a bioinformatic approach to systematically identify and classify CTL escape mutations in the *gag* gene of HIV-1 proviral sequences from a Kenyan sex worker cohort and correlated them with host HLA genotype and HIV disease markers.

The *gag* region encodes the primary structural proteins of the virus and contains many well-characterized immunodominant CTL epitopes (21). The Gag precursor protein, $Pr55^{gag}$, is cleaved by HIV-1 protease, resulting in mature p17 (matrix), p24 (capsid), p7 (nucleocapsid), p6, and two spacer peptides,

* Corresponding author. Mailing address: National Microbiology Laboratory, 1015 Arlington St., Winnipeg, Manitoba R3E 3R2, Canada. Phone: (204) 789-6074. Fax: (204) 789-2018. E-mail: ma_luo@phac-aspc.gc.ca.

† H.O.P., M.G.M., and R.E.C. made equal contributions to the work.

▽ Published ahead of print on 5 December 2007.

p1 and p2 (22). The various *gag* proteins play important roles in viral assembly (64, 85), budding (13, 84), maturation (25), stabilizing structural integrity (36), and infectivity (83). During the early stages of infection, the proteins of *gag* are involved in localization of the preintegration complex (34, 70), and therefore the conservation of the *gag* domains is critical for proper viral function.

Positively selected amino acids were first identified by examining the individual proteins of *gag* from more than 1,000 proviral *gag* sequences from the HIV-1-positive women of a cohort from Nairobi, Kenya. Positive selection sites were then classified as beneficial or detrimental to the virus by correlation with the patients' HLA types and CD4 counts. The positively selected amino acids were further correlated with the predicted proteasomal cleavage sites (using NetChop 3.0) and transport efficiency (using a TAP affinity algorithm). This allowed us to classify the positively selected amino acids that arose and in what manner they affect each step of peptide presentation. It explains why different HLA alleles are associated with dissimilar disease outcomes and is an efficient approach to identifying and classifying escape mutations in HIV-1 *gag*. This information can be used in HIV vaccine development.

MATERIALS AND METHODS

Study cohort. The patients were antiretroviral treatment-naïve HIV-1-positive adult women, at various stages of disease progression, enrolled in the Pumwani Sex Worker Cohort in Nairobi, Kenya. This study has been approved by the Ethics Committee of the University of Manitoba and the Ethics and Research Committee of Kenyatta National Hospital. Informed consent was obtained from all women enrolled in the study.

HLA sequencing and typing. Genomic DNA was isolated from 468 HIV-1-positive women enrolled in the Pumwani Sex Worker Cohort. HLA class I typing was conducted by amplifying HLA-A, -B and -C genes with gene-specific primers. The amplified PCR products were purified and sequenced using the ABI 3100 Genetic Analyzer. The class I genes were typed using the CodonExpress software package, which was developed based on taxonomy-based sequence analysis (55).

***gag* PCR and sequencing.** Proviral DNA was isolated from HIV-1-positive women. Nested PCR amplification was used to amplify *gag* genes. PCR amplification was confirmed using 1% agarose gel electrophoresis. PCR products were purified using the Multiscreen$_{HTS}$ PCR plate (Millipore Corp.). BigDye Terminator v3.1 was used to sequence *gag* genes with specific primers. The sequencing products were purified by ethanol-sodium acetate precipitation. Purified sequencing products were analyzed with an ABI 3100 Genetic Analyzer (Applied Biosystems). Nucleotide sequences were assembled and edited with Sequencher 4.5 (Genecodes Corp.). Samples with unsuccessful sequencing results due to heterogeneous quasispecies sequence were gel purified and cloned using TOPO TA cloning kit (Invitrogen). Multiple clones were sequenced as described above.

Phylogenetic analysis. Phylogenetic analysis with MEGA (Molecular Evolutionary Genetics Analysis) v3.1 was used to classify viral subtypes. All of the sequences were aligned using Clustal W (49), along with reference sequences obtained from the Los Alamos HIV database (47). Phylogenetic trees were constructed using neighbor-joining algorithms with bootstrap testing of 1,000 replicates. RIP (Intersubtype Recombination Analysis) v2.0 (47) was used to identify intersubtype recombinations.

Shannon's entropy was used to score the sequence variability in Gag protein alignments using the procedure described by Korber et al. (48). The score considers both the number of amino acid variants and their frequencies for each position, providing a quantitative measure for comparisons of *gag* sequences.

Positive selection analysis. QUASI, a selection-mapping algorithm (80), was used to identify the positively selected amino acids of viral proteins. This selection-mapping program identifies replacement mutations that are overabundant compared to silent mutations at each codon, recognizing them as positively selected.

Proteasomal cleavage prediction. NetChop C-term 3.0 calculated cleavage values for Gag residues and identified potential proteasomal cleavage sites (38, 43). A low cleavage value indicates a low probability of proteasomal cleavage; a high value suggests a high probability of cleavage.

TAP affinity prediction. Prediction of the TAP affinity, and therefore the TAP transport efficiency, was performed using the consensus scoring methods described by Peters et al. (74). The TAP affinity score is the sum of the matrix elements of the C terminus and three N-terminal residues, for any arbitrary length, represented by log 50% inhibitory concentration (IC_{50}) values. This equation optimally applies to nonamers; however, it was also highly correlated to peptides with 10 to 18 amino acids. A low TAP score corresponds to a peptide well-suited for TAP binding, and a high TAP score corresponds to low TAP affinity.

HLA and CD4 correlations. All statistical analyses were done using SPSS 11.0. To reduce the number of HLA correlations, a Z_{max} test was done for each positively selected residue to perform a global test of association using a Web-based program (http://www.msbi.nl/ScoreTest/) (17, 78). At each positively selected residue, all three HLA class I loci were separately tested. This statistic is designed to detect departures from the multinomial assumptions caused by the clustering of the observations in one or a few categories (50). A corresponding *P* value was estimated by simulation (10,000); however, this could only detect *P* values above 1.00×10^{-4}. The significant results were then followed up with individual allelic assessments, in which the Fisher's exact test was used in cases where an expected count of any cell in a two-by-two table was less than 10, and all other cases employed Pearson's chi-square test. Comparison of mean CD4 counts between positively selected amino acids and the consensus was conducted using an independent samples *t* test.

Nucleotide sequence accession numbers. All p17, p24, p7, and p6 sequences analyzed in this study were submitted to the National Center for Biotechnology Information GenBank database and can be found under accession numbers EF160141 to EF164802. The p1 nucleotide sequences was also determined (data not shown).

TABLE 1. Study population HLA distribution

HLA-A alleles (frequency [%])	HLA-B alleles (frequency [%])	HLA-C alleles (frequency [%])
A*7401 (9.12)	B*5802 (10.62)	Cw*0602 (17.17)
A*0201 (8.48)	B*1503 (9.87)	Cw*0401 (12.98)
A*3001 (8.37)	B*4201 (7.73)	Cw*0701 (12.02)
A*3002 (7.83)	B*5301 (7.40)	Cw*1701 (10.62)
A*2301 (6.87)	B*5801 (5.47)	Cw*0202 (9.01)
A*6802 (6.76)	B*4501 (4.94)	Cw*1801 (6.44)
A*0202 (6.01)	B*5703 (4.40)	Cw*1601 (5.58)
A*6601 (5.47)	B*1510 (4.29)	Cw*0304 (5.47)
A*0101 (5.15)	B*8101 (4.29)	Cw*0802 (4.72)
A*0301 (4.72)	B*0702 (4.18)	Cw*0702 (3.11)
A*2902 (3.76)	B*4901 (3.97)	Cw*0302 (1.72)
A*3601 (2.90)	B*0801 (2.47)	Cw*0407 (1.72)
A*0205 (2.36)	B*4403 (2.47)	Cw*0704 (1.72)
A*2402 (2.36)	B*1801 (2.36)	Cw*0804 (1.39)
A*3402 (2.36)	B*3501 (2.25)	Cw*1505 (1.07)
A*3201 (1.72)	B*1302 (1.93)	Other[a] (5.26)
A*6801 (1.50)	B*1406 (1.82)	
A*3004 (1.18)	B*4415 (1.61)	
A*3009 (1.07)	B*5702 (1.39)	
Other[a] (12.01)	B*1402 (1.18)	
	B*5101 (1.18)	
	B*3910 (1.07)	
	Other[a] (13.11)	

[a] Frequencies of HLA alleles that were below 1.0% were grouped into the "Other" category.

RESULTS

HLA serotype and allele distribution. All 468 HIV-1-positive women were typed for their HLA-A, -B, and -C alleles. Nineteen HLA-A, 22 HLA-B, and 15 HLA-C alleles were present at a frequency greater than 1%, while rare alleles altogether were present at 12.01% ($n = 28$), 13.11% ($n = 51$), and 5.26% ($n = 14$), respectively. Refer to Table 1 for

TABLE 2. Clade distribution of *gag* proviral sequences

Clade	Frequency (%) of indicated protein in clade			
	p17	p24	p7	p6
A1	71.1	64.9	67.1	62.9
D	13.6	22.1	22.8	20.4
C	6.3	5.4	6.1	3.1
Other recombinant subtypes	8.7	7.6	4.0	13.6

TABLE 3. Shannon's average entropy of *gag* for clades A1 and D

Gag	Avg entropy for clade:	
	A1	D
p17	0.0963	0.0966
p24	0.0443	0.0357
p7	0.0617	0.0764
p1	0.1025	0.1025
p6	0.1682	0.1200

the specific distribution of HLA alleles within this subpopulation.

Phylogenetic analysis and sequence variability in HIV-1 *gag*. A total of 1,552, 1,346, 780, 1,224, and 984 sequences were generated from p17, p24, p7, p1, and p6, respectively, from 468 patients. Phylogenetic analysis, using MEGA 3.1, classified the viral subtypes based on clustering patterns with known reference sequences. To determine recombination, classification was conducted separately for each protein. The cohort is predominantly infected by clade A1 at 71%, 65%, 67%, and 63%, followed by clade D at 14%, 22%, 23%, and 20% for p17, p24, p7, and p6, respectively (Table 2). The p1 region is too short (16 residues) to use in phylogenetics with a reasonable degree of certainty and so was omitted from this analysis. Although clade C viruses are prevalent in the southern region of sub-Saharan Africa (9, 57, 71), in this East African population they are identified in only 10% of the viral population. Recombinant subtypes ranged from 4 to 13.6% in each individual Gag protein (Table 2). For statistical purposes, this study concentrated on subtypes A1 and D for each of the Gag proteins.

Viral escape mutations can occur at each step of antigen presentation, including antigen processing, antigen presentation, and T-cell receptor (TCR) binding. To systematically investigate viral escape mutations of HIV-1 in the Pumwani Sex Worker Cohort, an approach that identifies and classifies escape mutations following each step of antigen processing and presentation was adopted.

The viral protein sequence variability was first used to identify regions within *gag* that are indicative of viral evolution. Shannon's entropy score is a commonly used tool for establishing sequence variability by quantitatively measuring the presence of variants while filtering out random genetic drift (79). A high entropy score implies that many variants are present at significant frequencies, while a low score implies relative conservation. A comparison of average entropy for each *gag* gene product using an independent sample *t* test (p17, p24, p7, p1, and p6 individually tested against the rest of the Gag region) showed that p24 has the lowest score for both the A1 (2.4-fold less; $P = 2.98 \times 10^{-8}$) and D (2.8-fold less; $P = 3.95 \times 10^{-10}$) clades. This confirms the relative conservation of p24 (Table 3). This observation suggests that p24's structural constraints limit its diversity. Further analysis of p7-p1-p6 showed higher average entropy for both A1 and D clades (p7 < p1 < p6), implying that fewer structural constraints exist in p1 and p6 compared with the structural components of *gag* (p17, p24, and p7).

In HIV infection the persistent generation of variants in immune epitopes ultimately leads to a reduction of immune control. However, not all escape mutations are advantageous to the virus, as some can severely hinder viral fitness (33). Investigation of individual residue entropies across Gag reveals that regions in epitopes tend to occupy areas of lower entropy ($P = 0.013$, independent samples *t* test) (Fig. 1 and 2), an observation consistent with the findings of Yusim et al. (88). Amino acid variants in CTL epitopes often hinder or abolish the CTL response (18) and can point to locations of potential immune escape.

Identification of positively selected amino acids using QUASI analysis. Positively selected amino acids identified by QUASI were used to characterize the selection landscape of p17, p24, p7, p1, and p6 (Fig. 3). Because of the significant genetic distance and complications of interclade variability between clades A1 and D, we conducted QUASI analyses separately for each clade. Clade A1 and D subtypes differed in the number of positively selected sites, number amino acid variants, and frequencies of variants. A positive selection map across *gag* was generated (Fig. 3). The positively selected amino acids were then analyzed for selection pressures that may drive viral evolution at the various avenues of escape.

Figure 4 shows a comparison between conserved and nonconserved mutations in the individual Gag protein products from both clades A1 and D. Both types of mutations were classified using the BLOSUM62 amino acid substitution matrix (35). Since p17 and p24 have more structural constraints than the other Gag proteins, fewer nonconserved positively selected mutations were observed. In contrast, p6 contains the most nonconserved mutations in clades A1 ($P = 7.46 \times 10^{-8}$) and D ($P = 3.57 \times 10^{-4}$).

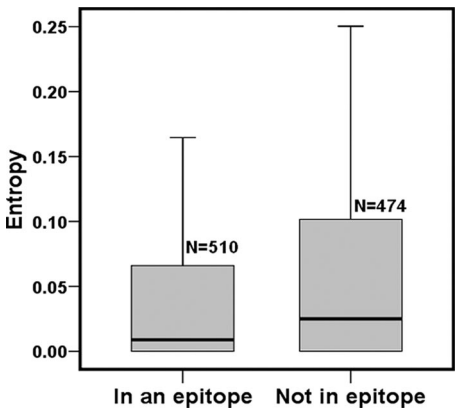

FIG. 1. Comparison of individual residue entropies between those areas in Gag that lie in known epitopes and those that do not ($P = 0.013$, independent sample *t* test).

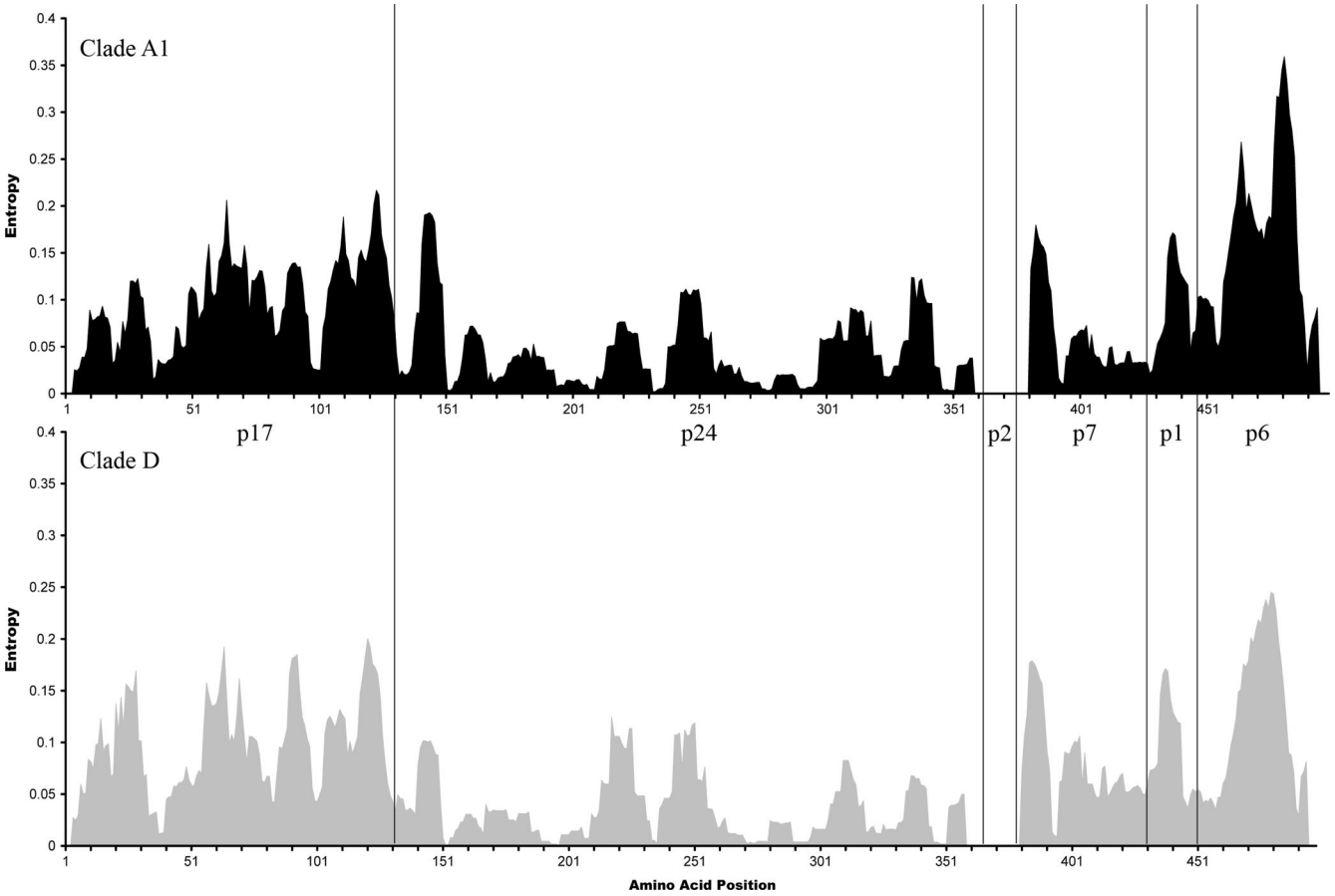

FIG. 2. Calculation of the average Shannon's entropy value across a 9-residue window at each position of Gag. (A) Clade A1 plots (black); (B) clade D (grey). p2 was not analyzed.

Determining proteasomal escape mutations. NetChop 3.0 at a threshold value of 0.5 was utilized to assess selections that may abrogate viral peptide processing for HLA class I presentation. However, this prediction algorithm only determines C-terminal cleavage sites, because the determination of N-terminal cleavage sites is more complicated (43). Positively selected mutations that occur on the C-terminal cleavage sites may abolish proteasomal processing, and any such mutations that flank the C-terminal cleavage site and occur within 14 amino acid residues may also affect cleavage. Therefore, positively selected mutations that occur at the C terminus of the epitope and residues within, or flanking the epitope, were analyzed. Positively selected mutations affecting the proteasomal cleavage are marked in Fig. 3.

In p17 of subtype A1, the HLA-A3-restricted RK9 epitope contained a positive selection at the C-terminal anchor residue K28Q (Fig. 3). Conserved Lys-to-Gln mutations at this site abolish proteasomal cleavage, as indicated by the dramatically reduced NetChop score (from 0.625 to 0.093). The K28Q mutation at the C-terminal proteasome cleavage site is not affected by other mutations within a 14-residue sequence range. This suggests that the K28Q mutation is a proteasome escape variant. Patients carrying the K28Q mutation tend to have a lower CD4 count; however, the difference is not significant (data not shown). This observation is consistent with a previous study that suggested the K28Q mutation might impair HLA-A3-restricted epitope processing and reduce binding (2).

The extended peptide generated from proteasomal cleavage requires N-terminal trimming in the ER, an essential step for some epitopes in peptide processing (8, 77). For instance, in p24, the positively selected amino acid A146P flanking the immunodominant HLA-B57-restricted epitope IW9 (p24, 147 to 155) was found previously to prevent N-terminal trimming by the ER aminopeptidase I (15). In this Kenyan HIV-1-infected population, the A146P mutation in clade D was associated with a lower mean CD4 count (from 347.69 to 283.52 cells/ml), although it was not statistically significant ($P = 0.20$). Another positively selected amino acid, I147L, located within the same epitope was significantly associated with a decrease in the mean CD4 count (from 356.36 to 283.52 cells/ml) ($P = 0.014$; see Table 7, below). Goulder et al. (29) found that A146P and I147L occurred at the same time in HLA-B57 progressors. A146P and I147L simultaneously occurred in ~11% of the Kenyan patients and were associated with HLA-B*5703 ($P = 4.41 \times 10^{-5}$ and 4.02×10^{-2}, respectively).

Analyses of Kenyan p7 sequences revealed no positively selected mutations that can alter proteasomal cleavage sites within the two C-terminal residues containing the conserved zinc fingers and the basic linker peptide (10, 76). Kenyan A1 clade viruses only have changes to the predicted cleavage sites

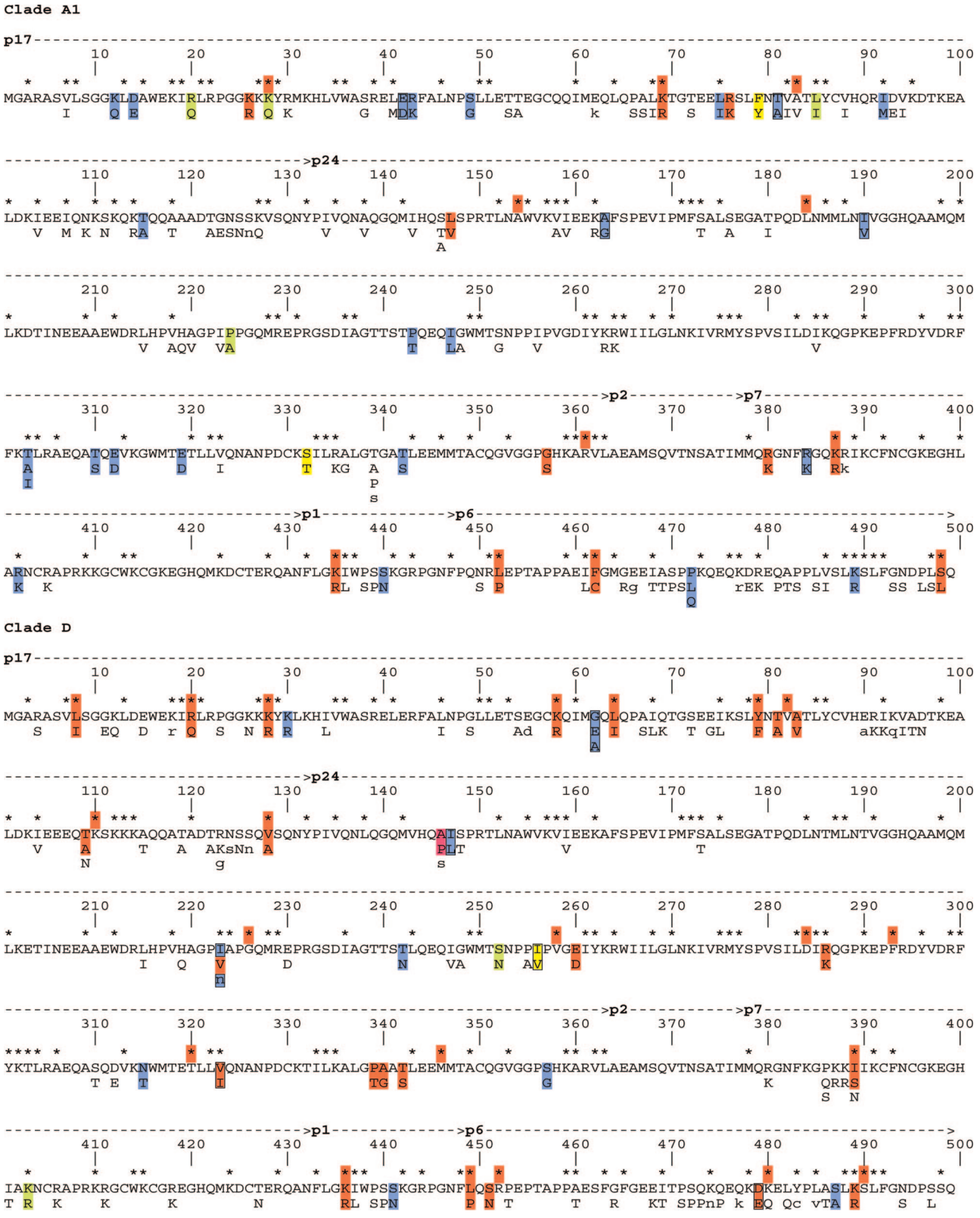

FIG. 3. A map of positively selected mutations across *gag* that were generated by QUASI for clade A1 and D viruses. The consensus sequence is shown as a single line of residues with positively (upper case) and neutral (lower case) selected residues shown underneath the consensus for each site. Predicted proteasomal cleavage sites of the consensus sequence are shown as asterisks at an IC_{50} threshold score of 0.5. Mutations that affect proteasomal cleavage are shown in red, and those cleavage sites that are abolished through mutations are shown as red asterisks. Mutations that affect predicted TAP transport are shown in green; mutations that affect N-terminal trimming are in pink; mutations that may result in reduction of HLA binding are in blue; mutations that may reduce TCR recognition are in yellow. Mutations that correlate with differences in CD4 counts are boxed.

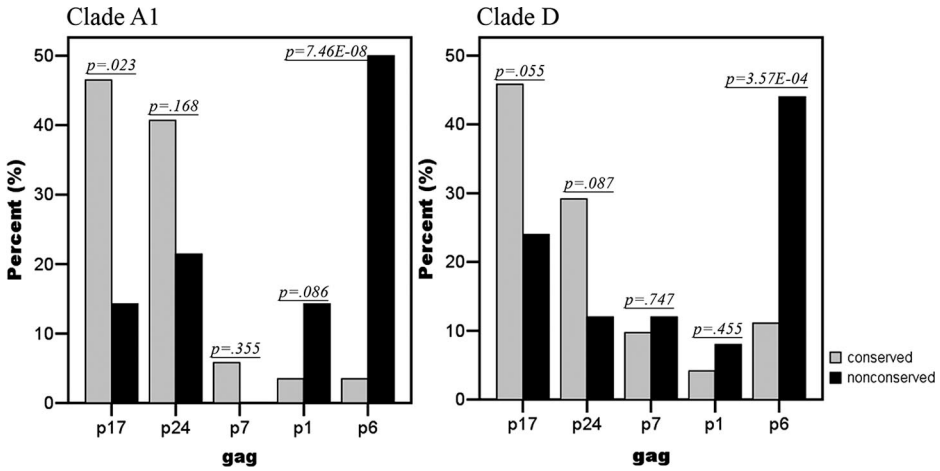

FIG. 4. Comparison of conserved and nonconserved positively selected mutations in p17, p24, p1, p7, and p6 from both clades A1 and D. Conserved (grey) and nonconserved (black) mutations are shown with their respective *P* values. Classifications of amino acid substitution were based on the BLOSUM62 amino acid substitution matrix (35).

within the first 12 residues of the protein. The R380K mutation is predicted to result in a removal of one of these sites at residue 387 with the concomitant formation of a new site at residue 380. Similarly, among the D clade viruses, positively selected mutations from residues 386 through 389 resulted in a number of proteasomal cleavage sites either being created or destroyed. However, as was the case with A1 clade viruses, these mutations all occurred within the first 12 residues. There are no known epitopes that have been identified in this region of p7. The I401T and R418K mutations in the D clade viruses resulted in the creation of cleavage sites at residues 397 and 418, respectively; however, these new sites occur outside of documented epitopes and are not likely to influence the processing of the known epitopes.

Determining TAP escape mutations. A TAP affinity algorithm was used to determine TAP-peptide binding log IC_{50} scores (74) (Table 4) and predict TAP escape mutations. A higher log IC_{50} value indicates a lower TAP-peptide binding affinity. Positive selection was observed in the A3-restricted epitope RK9 at sites 1 and 9 (R20Q and K28Q, respectively). The change from RK9 to RQ9 resulted in an increase of the log IC_{50} value from −2.12 nM to −1.55 nM, while RK9 to QK9 resulted in a greater increase of the log IC_{50} value from −2.12 nM to −0.31. These mutations show that TAP binding affinity is affected by certain mutational variants and suggest that TAP escape mutations may occur.

TABLE 4. TAP affinity scores for RK9, SL9, MV9, and LARK10 of *gag*

Clade	*gag*	Epitope	Sequence[a]	Log IC_{50} (nM)[b]
A1	p17	RK9	**R**LRPGGKK**K**	−2.12
			RLRPGGKK**Q**	−1.55
			QLRPGGKK**K**	−0.31
			QLRPGGKK**Q**	0.26
A1		SL9	SL**F**NTVATL	−2.46
			SL**Y**NTVATL	−3.21
D	p24	MV9	MT**S**NPPIPV	−0.61
			MT**N**NPPIPV	0.29
D	p7	LARK10	**LAR**NCRAPRK	−1.62
			IAKNCRAPRK	0.09
			IARNCRAPRK	−1.23

[a] The consensus sequence for each epitope is listed first, followed by the mutant sequence(s) found. The sites of variants are shown in boldface, with the corresponding TAP score (log IC_{50}).

[b] The log IC_{50} values were calculated by summing the values of the consensus scoring matrix described by Peters et al. (74) of the three N-terminal residues and the single C-terminal residue. See reference 74 for a complete description.

Positive selection in p24 that was indicative of a TAP escape mutation was also observed. In clade D, a change in amino acid residue 252 from Ser to Asn within the HLA-A*0201-restricted epitope increased the log IC_{50} value from −0.61 to 0.29 nM, implying a decrease in binding affinity.

The p7 of the D clade contains a positively selected mutation that seems to affect TAP binding. The HIV-1 database lists the region $_{401}$LARNCRAPRK$_{410}$ (LARK10) as an epitope presented by HLA-A3. In this Kenyan cohort, the D clade consensus sequence at this site is $_{401}$IAKNCRAPRK$_{410}$ (IAKK10). The mutation, K403R, correlates with HLA-A*0301 ($P = 8.07 \times 10^{-4}$), indicating that this sequence is also potentially under selection pressure by HLA-A3. The TAP binding score (log IC_{50}) for the LARK10 epitope is −1.62 nM, a score that indicates excellent affinity for TAP. IAKK10 increased the TAP score by almost 2 logs to 0.09, indicating a substantial decrease in TAP affinity. This suggests that adaptation has occurred at this site for clade D virus in this cohort.

Determining HLA escape mutations. Mutations in the anchor positions may lead to a reduction in HLA binding and abolish peptide presentation (42). This was shown by correlating HIV-1 positively selected mutations identified using QUASI with the host HLA alleles. To reduce the number of HLA correlations, a Z_{max} test was first used to globally detect associated HLA alleles at each positively selected residue (50). For each significant Z_{max} test, a cross-tabulation (Fisher's exact test and Pearson's chi-square test) was then used to correlate the positively selected amino acids with the host HLA alleles (Tables 5 and 6; note that p1 was not included in phylogenetic analysis due to its short length, and thus the p1 correlations

TABLE 5. HLA correlations to positively selected amino acids in clade A1 Gag proteins

Gag protein	Mutation	Locus	Z_{max} test		HLA allele	Correlation[b]	Odds ratio (95% CI)[c]	Odds ratio *P* value[d]	Reference(s)[e]
			Score	*P* value[a]					
p17	K12Q	HLA-A	40.97	1.51×10^{-2}	A*7401	+	7.95 (5.95–10.62)	6.37×10^{-7}	Undocumented
	D14E	HLA-A	24.21	1.88×10^{-2}	A*7401	+	10.46 (3.75–29.17)	9.05×10^{-5}	Undocumented
	R20Q	HLA-A	35.03	2.80×10^{-3}	A*7401	+	11.83 (4.31–32.46)	1.71×10^{-6}	Undocumented
	K26R	HLA-C	12.84	1.27×10^{-2}	Cw*0602	+	3.73 (1.88–7.40)	**8.81×10^{-5}**	Undocumented
	K28Q	HLA-A	57.60	$<1.00 \times 10^{-4}$	A*0301	+	19.58 (8.06–47.55)	3.97×10^{-12}	29, 31, 65
					A*3001	+	4.27 (2.20–8.27)	3.25×10^{-5}	29
	E42D	HLA-B	33.12	$<1.00 \times 10^{-4}$	B*3502	+	10.82 (2.33–50.35)	4.74×10^{-3}	Undocumented
		HLA-C	44.02	$<1.00 \times 10^{-4}$	Cw*0407	+	3.99 (1.14–13.92)	4.28×10^{-2}	Undocumented
	R43K	HLA-C	17.96	1.00×10^{-2}	Cw*0304	+	8.20 (2.78–24.18)	4.66×10^{-4}	Undocumented
	S49G	HLA-B	15.90	1.90×10^{-3}	B*1402	+	34.74 (1.96–614.38)	2.78×10^{-4}	Undocumented
	L75I	HLA-A	22.24	$<1.00 \times 10^{-4}$	A*0202	+	5.62 (2.67–11.83)	4.59×10^{-6}	40
	F79Y	HLA-A	17.22	5.00×10^{-4}	A*0101	−	0.16 (0.06–0.42)	**4.02×10^{-5}**	40, 62
					A*0202	+	5.94 (2.39–14.79)	**2.26×10^{-5}**	29, 40
					A*3002	−	0.23 (0.11–0.51)	**8.38×10^{-5}**	28, 40, 69
					A*3601	−	0.10 (0.02–0.42)	**1.55×10^{-4}**	Undocumented
	T81A	HLA-A	32.93	5.50×10^{-3}	A*0201	−	0.06 (0.00–0.96)	1.50×10^{-3}	3, 14, 40, 72
	V82I	HLA-A	16.32	3.38×10^{-2}	A*0240	+	19.37 (3.08–121.82)	4.11×10^{-3}	Undocumented
	V88I	HLA-C	17.83	3.57×10^{-2}	Cw*1601	+	11.13 (2.84–43.58)	1.30×10^{-3}	Undocumented
	I92M	HLA-A	36.81	1.83×10^{-2}	A*0204	+	42.00 (5.71–308.92)	3.87×10^{-3}	Undocumented
	D93E	HLA-C	22.52	1.00×10^{-4}	Cw*0602	−	0.20 (0.11–0.37)	**6.18×10^{-8}**	Undocumented
	T115A	HLA-A	40.83	$<1.00 \times 10^{-4}$	A*3001	+	8.66 (4.17–17.9)	1.87×10^{-8}	Undocumented
p24	A163G	HLA-B	38.05	$<1.00 \times 10^{-4}$	B*5703	+	12.71 (4.97–32.49)	3.21×10^{-8}	29, 32
	I190V	HLA-B	14.77	4.38×10^{-2}	B*5802	+	4.06 (1.95–8.44)	4.73×10^{-4}	26
		HLA-C	27.24	2.00×10^{-4}	Cw*0701	−	0.33 (0.10–1.11)	4.14×10^{-2}	Undocumented
	P243T	HLA-B	17.12	1.76×10^{-2}	B*5702	+	7.59 (1.92–29.95)	9.17×10^{-3}	6, 61
					B*5703	+	4.19 (1.58–11.10)	7.42×10^{-3}	6
	I247L	HLA-B	17.52	1.88×10^{-2}	B*5702	+	10.95 (2.78–43.08)	1.47×10^{-3}	30, 54, 61
	T303A/I	HLA-C	21.91	2.00×10^{-3}	Cw*0304	+	5.75 (2.60–12.70)	3.94×10^{-5}	68
	T310S	HLA-B	28.96	$<1.00 \times 10^{-4}$	B*4415	+	10.19 (2.45–42.31)	1.22×10^{-3}	65
					B*4901	+	10.58 (2.55–43.85)	2.47×10^{-6}	Undocumented
	E312D	HLA-B	23.37	1.00×10^{-4}	B*4901	+	8.09 (3.01–21.70)	1.38×10^{-5}	Undocumented
					B*5802	−	0.31 (0.13–0.71)	**3.90×10^{-3}**	65
		HLA-C	13.60	7.70×10^{-3}	Cw*0701	+	3.18 (1.69–5.99)	**1.29×10^{-4}**	Undocumented
	E319D	HLA-B	18.93	2.10×10^{-3}	B*4501	+	5.73 (2.49–13.21)	6.65×10^{-5}	Undocumented
	T342S	HLA-C	19.37	1.60×10^{-3}	Cw*0804	+	11.97 (2.99–48.01)	3.59×10^{-4}	Undocumented
	G357S	HLA-C	21.99	1.00×10^{-4}	Cw*1601	+	6.07 (2.78–13.27)	5.63×10^{-6}	Undocumented
p7	K387R	HLA-A	18.22	4.80×10^{-3}	A*7401	+	3.32 (1.11–9.95)	1.57×10^{-4}	Undocumented
	R402K	HLA-B	12.98	3.72×10^{-2}	B*1405	+	13.85 (1.41–136.11)	2.14×10^{-2}	Undocumented
p1	K435R	HLA-B	15.71	1.01×10^{-2}	B*1302	+	10.09 (2.54–40.02)	7.83×10^{-4}	Undocumented
	S440N	HLA-A	12.24	1.86×10^{-2}	A*7401	+	3.12 (1.66–5.87)	**2.46×10^{-4}**	Undocumented
p6	P472LQ	HLA-C	10.86	2.79×10^{-2}	Cw*0202	+	8.26 (2.00–34.08)	3.39×10^{-3}	Undocumented
	K489R	HLA-C	43.49	5.80×10^{-3}	Cw*0304	+	8.93 (1.60–49.91)	3.93×10^{-2}	Undocumented
					Cw*1701	+	5.45 (1.67–17.82)	5.79×10^{-3}	Undocumented

[a] The *P* value for the Z_{max} test was estimated by 10,000 simulations. The Web-based program cannot detect *P* values below 1.00×10^{-4}.
[b] Positive (+) and negative (−) HLA correlations are shown.
[c] CI, confidence interval.
[d] *P* values shown in bold denote results obtained from a Pearson chi-square analysis.
[e] A reference that is "undocumented" corresponds to a potential novel epitope.

presented in Tables 5 and 6 are the same). This results in a total of 2,338 tests for clades A1 and D. If several sequences were available for a single patient, those sequences were consolidated and each residue was examined for the presence of a positively selected mutation. This ensured that each host (and their corresponding HLA allele) was only counted once. This analysis makes it possible to identify potential epitopes for rare HLA alleles and undocumented epitopes. Although HLA alleles are codominantly expressed, we assumed a dominant genotype model in the analysis, since one copy of a specific HLA allele is enough to exert a sufficiently strong selective pressure on the virus (32, 44). This means that an escape variant will occur whether there are one or two copies of a specific HLA allele. Overall, 36 positive and 7 negative HLA correlations were observed in clade A1, while 16 positive and 1 negative HLA correlation were found in clade D. Eighty-three percent of the identified correlations with *P* values less than 0.01 have been documented by previous studies. Thirty-five of the correlations have not been documented by previous studies, and among them, 80% have *P* values less than 0.01. While controlling for multiple comparisons is important, it should be understood that this analysis was used as a preliminary screening for

TABLE 6. HLA correlations to positively selected amino acids in clade D Gag proteins

Gag protein	Mutation	Locus	Z_{max} test[a]		HLA allele	Correlation[b]	Odds ratio (95% CI)[c]	*P* value[d]	Reference(s)[e]
			Score	*P* value					
p17	K28R	HLA-A	11.45	8.40×10^{-3}	A*3001	+	8.68 (2.33–32.37)	1.02×10^{-3}	29
	K30R	HLA-A	9.88	2.21×10^{-2}	A*2301	+	9.00 (2.13–37.96)	1.96×10^{-3}	44, 45
					A*2402	+	18.72 (0.93–377.67)	2.37×10^{-2}	38
	G62E/A	HLA-B	13.02	1.50×10^{-3}	B*5301	+	24.63 (2.92–207.63)	2.02×10^{-4}	Undocumented
p24	A146P	HLA-B	25.65	5.00×10^{-4}	B*5703	+	47.50 (5.26–428.91)	4.41×10^{-5}	66
	I147L	HLA-B	10.09	2.14×10^{-2}	B*5703	+	5.61 (1.03–30.39)	4.02×10^{-2}	5, 40
	T242N	HLA-B	33.02	1.00×10^{-4}	B*5703	+	47.50 (5.26–428.91)	7.30×10^{-4}	6
	P255A	HLA-A	18.27	2.36×10^{-2}	A*0201	+	8.95 (1.81–44.29)	2.33×10^{-2}	58
	N315T	HLA-B	53.97	3.00×10^{-4}	B*4415	+	217.00 (6.11–5,056.94)	2.76×10^{-3}	1, 2, 21, 65
	V323I	HLA-B	17.91	2.22×10^{-2}	B*5101	+	26.50 (2.94–239.07)	1.33×10^{-2}	Undocumented
	S357G	HLA-B	13.49	1.10×10^{-3}	B*0702	+	21.21 (2.61–172.68)	7.67×10^{-5}	33, 45
p7	K403R	HLA-A	10.84	4.60×10^{-3}	A*0301	+	18.36 (2.20–153.06)	8.07×10^{-4}	12
					A*3001	+	4.32 (1.21–15.43)	2.89×10^{-2}	Undocumented
					A*6802	−	0.10 (0.13–0.82)	1.26×10^{-2}	Undocumented
p1	K436R	HLA-B	15.71	1.01×10^{-2}	B*1302	+	10.09 (2.54–40.02)	7.83×10^{-4}	Undocumented
	S441N	HLA-A	12.24	1.86×10^{-2}	A*7401	+	3.12 (1.66–5.87)	$\mathbf{2.46 \times 10^{-4}}$	Undocumented
p6	S487A	HLA-A	29.76	2.40×10^{-3}	A*6802	+	22.50 (4.07–124.39)	1.97×10^{-4}	Undocumented

[a] The *P* value for the Z_{max} test was estimated by 10,000 simulations. The Web-based program cannot detect *P* values below 1.00×10^{-4}.
[b] Positive (+) and negative (−) HLA correlations are shown.
[c] CI, confidence interval.
[d] *P* values shown in bold denote results obtained from a Pearson chi-square analysis.
[e] A reference that is "undocumented" corresponds to a potential novel epitope.

potential HLA epitopes, and for this reason, all results have been included and the *P* values have not been corrected, and we acknowledge the possibility of type I error. If one desires, different methods of multiple comparison corrections can be applied directly from Tables 5 and 6 with the total number of tests provided above.

Figure 5 shows an epitope map based on previous studies, as listed by Frahm et al. (21) and the Los Alamos Database (47). Clusters of positive selections within defined epitopes are evident in p24; however, selections are more uniformly distributed among p17 and p6. There is also evidence that epitopes restricted by several HLA class I molecules cluster in some regions, suggesting that some parts of individual proteins are more immunogenic than others.

K28Q, in p17 A1, is located within the β-sheet basic domain between helices 1 and 2. It is significantly associated with the

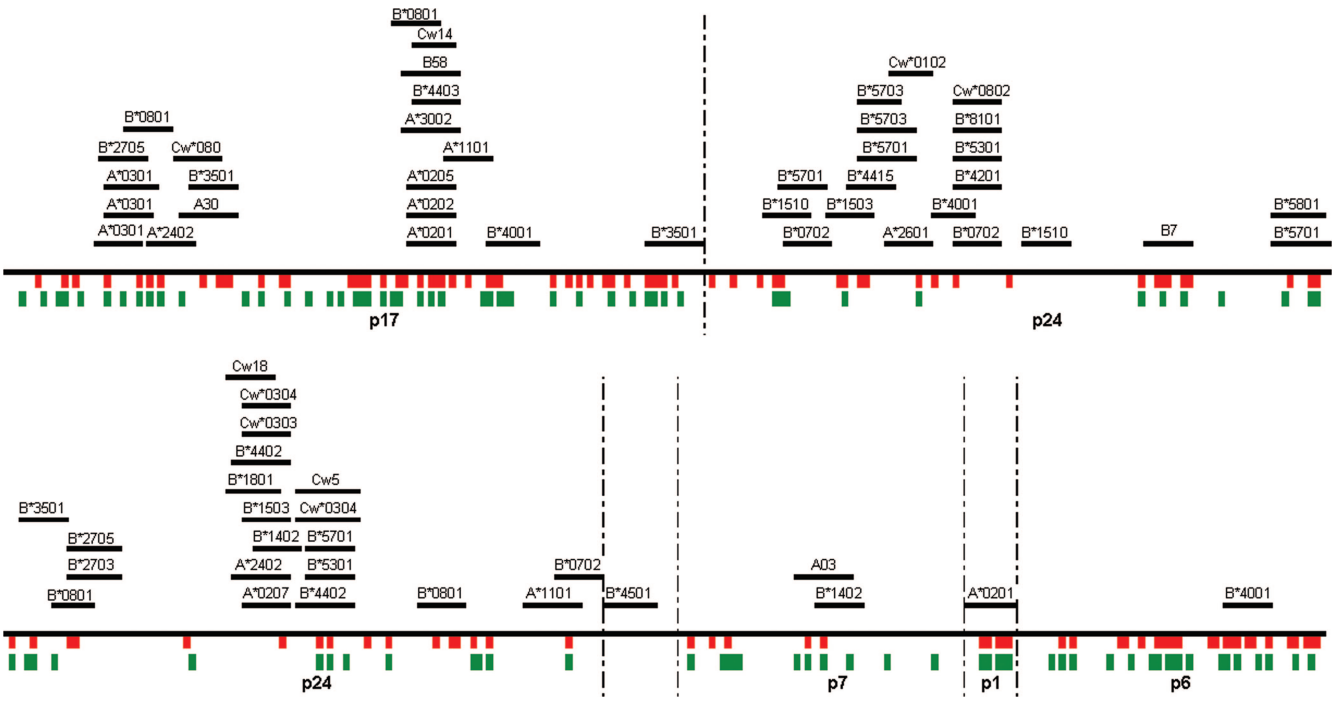

FIG. 5. Epitope map of Gag, based on the optimal HIV-1 CTL epitopes defined by previous studies found in Table I-A-1 of Frahm et al. 2005 (21). The locations of the positively selected amino acids in the Kenyan cohort are shown in red (clade A1) and green (clade D).

TABLE 7. Comparison of mean CD4 counts between HIV-1-positive individuals with consensus versus positively selected amino acids

Clade	Gag protein	Mutation	ΔCD4[a]	*P* value[b]	Escape
A	p17	E42D	−75.25	1.16×10^{-2}	HLA
		T81A	−89.08	7.30×10^{-3}	HLA
	p24	A163G	88.04	3.77×10^{-2}	HLA
		I190V	−92.31	1.76×10^{-2}	HLA
	p7	R384K	−70.21	1.60×10^{-2}	HLA
D	p17	G62E/A	−140.79	1.13×10^{-3}	HLA
	p24	I147L	−72.83	1.38×10^{-2}	HLA
		I223n	228.06	7.31×10^{-4}	HLA
		P255A	−167.70	1.92×10^{-3}	TCR
		V323I	−125.93	1.62×10^{-2}	Proteasome
	p6	D479E	−139.55	1.60×10^{-2}	Proteasome

[a] The ΔCD4 count was calculated as the difference in CD4 counts for the mutant versus consensus sequence.
[b] The *P* value was derived from an independent samples *t* test.

HLA-A*0301 genotype ($P = 3.97 \times 10^{-12}$) (Table 5), which is known to bind the epitopes $_{18}$KIRLRPGGK$_{26}$ (KK9) (39) and $_{20}$RLRPGGKKK$_{28}$ (RK9) (7, 87). Since K28Q is located in the C terminal of the epitope RK9, it is possible that this mutation may also result in a reduction in HLA-A*0301 recognition.

The HLA-A2-restricted epitope SL9 had several positively selected mutations, including F79Y and T81A in the A1 clade. Site F79Y had significant positive correlations with HLA-A*0202 ($P = 2.26 \times 10^{-5}$). This site also had significant negative associations with HLA-A*0101 ($P = 4.02 \times 10^{-5}$), A*3002 ($P = 8.38 \times 10^{-5}$), and A*3601 ($P = 1.55 \times 10^{-4}$) (Table 5). HLA-A*0101 ($_{71}$GSEELRSLY$_{79}$), HLA-A*0201 (SL9), and HLA-A*3002 ($_{76}$RSLYNTVATLY$_{86}$) epitopes were previously identified in this region (16). These negative correlations suggest that the consensus phenylalanine is an escape mutation that has become fixed in the population (52). The mutation T81A is also negatively associated with A*0201 ($P = 1.50 \times 10^{-3}$), which is correlated with a lower mean CD4 count ($P = 7.30 \times 10^{-3}$) (Table 7). Also, positively selected mutations flanking SL9, such as L75I and I92M, that are positively correlated to HLA-A2 (particularly A*0202, $P = 4.59 \times 10^{-6}$; A*0204, $P = 3.87 \times 10^{-3}$) are possible escape mutations that may prevent both peptide processing and recognition of the SL9 epitope.

The effect of HLA alleles associated with long-term nonprogression on HIV-1 evolution was examined by analyzing and classifying positive selection in Gag proteins of this Kenyan study population's HLA alleles, such as B*5701 and B*5703. These particular alleles have been strongly associated with slow progression to disease and thus theoretically would exert a strong selective pressure on the virus (53). Five immunodominant epitopes restricted by HLA-B57 have previously been located in the p24 region: $_{147}$ISPRTLNAW$_{155}$ (IS9), $_{162}$KAFSPEVIPMF$_{172}$ (KAF11), $_{162}$KAFSPEVI$_{169}$ (KAF8), $_{240}$TSTLQEQIAW$_{249}$ (TW10), and $_{308}$QASQEVKNW$_{316}$ (QW9) (52, 56, 61, 81). The positively selected mutation in clade A1, A163G at the P2 anchor position of the KAF11 and KAF8 epitopes, is highly correlated with patients expressing HLA-B*5703 ($P = 3.21 \times 10^{-8}$). This mutation is also associated with an increased mean CD4 count from 337.52 cells/ml to 425.56 cells/ml ($P = 0.038$) (Table 7). It has been shown that HLA-B*5703 can still present the A163G epitope variant efficiently (27). This could be a basis for the association of HLA-B*5703 with slower disease progression.

This method also identified viral sequences correlated with HLA alleles associated with rapid disease progression, such as HLA-B*5802 (44). HLA-B*5802 in this Kenyan population is highly correlated with the positively selected mutation at residue 190 from Ile to Val ($P = 4.73 \times 10^{-4}$) in subtype A1-infected individuals. This mutation correlates to a decrease in mean CD4 count from 364.04 to 271.73 CD4 cells/ml ($P = 0.018$), suggesting that patients harboring viral strains with V190 progress faster than those who harbor strains with I190. This mutation is also negatively correlated with HLA-Cw*0701 ($P = 1.94 \times 10^{-3}$), implying selective pressure toward consensus in its presence. This is expected, since Cw*0701 is associated with a reduced susceptibility to HIV infection (20).

Viral mutations that spare the anchoring residues of an epitope should not affect HLA binding; however, viral evasion might still occur due to impaired recognition of the HLA-peptide complex by the TCR (61). In this study population, the V82I mutation within the SL9 epitope restricted by HLA-A*0201 correlated with a decrease in mean CD4 count from 351.13 cells/ml to 277.48 cells/ml, although this result was not significant ($P = 0.142$). It has been reported in a longitudinal study that V82I escape mutations arose within 2 weeks of selection pressure from a Gag-specific CTL clone (86).

An indication of a loss of CTL recognition is seen in the p24 clade D sequences. The mutation occurs in amino acid 255, which is located in the P6 position within the HLA-A*0201-resticted epitope. This Pro-to-Ala mutation also correlates with a decrease in mean CD4 count from 334.03 to 116.33 CD4 cells/ml ($P = 0.002$), implying that this mutation results in faster disease progression.

DISCUSSION

HIV-1 evolution is a result of the combined effects of error-prone reverse transcriptase, viral fitness restrictions, and host selective pressure. Many selection algorithms have been developed to predict viral variants that portray selective advantages. In this study, we used QUASI to illustrate the selection landscape in *gag* by calculating the overabundance of replacement mutations relative to silent mutations. The selection map generated was used to study the potential effects that these positive selections have on various steps in antigen presentation. A brief summary of these findings can be found in Fig. 3.

In this report, an integrated approach to identify suspected CTL escape mutations in HIV-1 *gag* was taken. Although many studies have shown that HLA binding and CTL recognition contribute the most specificity in cellular-mediated immune responses, there has been increasing evidence that both proteasomal cleavage and TAP affinity may contribute as well (8, 63, 67, 73, 74, 82). Therefore, integrated predictions of C-terminal proteasomal cleavage, TAP affinity, and HLA class I associations were correlated with CD4 counts.

QUASI is similar to other selection algorithms, such as dN/dS and DataMonkey (75, 80), in identifying selections. However, our sample sizes were greater than 1,000 sequences; hence, QUASI was better suited, as other selection algorithms have a 100-sequence limitation. QUASI, like other programs,

only considers single-nucleotide mutations within the consensus codon; hence, it disregards variants in codons that contain two point mutations from consensus and classifies those as undefined mutations. This is problematic, since some undefined variants were overabundant. For example, in p24 clade A1 (Fig. 3), QUASI reported that residue 315 contains Gly (n = 715) as the consensus. However, in the same codon, there exists an abundance of double-point mutations coding for Asn (n = 243), but Asn is classified as an undefined variant by QUASI. If Asn were generated through a single-point mutation from the consensus, it would most likely be categorized by QUASI as positively selected. Considering that reverse transcriptase generates about one mutation out of 10 kb, two mutations occurring within a codon is highly improbable. It is possible that Asn did not arise from a mutation but exists as an established variant that coexists with Gly, as is shown in clade D (Fig. 3); Asn and Gly exist as the consensus and an undefined variant, respectively, at residue 315. Furthermore, in other clades, the consensus of amino acid 315 is Gly in clades F1 and G and Asn in clades B and C (47).

Proteasomal escape variants (those that abolish C-terminal cleavage sites upon mutations that occur in the C terminus of the epitope, within the epitope, or flanking the epitope) were identified. Strong internal cleavage sites that could destroy potential epitopes were not considered possible escape variants, since the QUASI output is generated from a population with different HLA-restricted epitopes. It may be possible, therefore, to observe multiple cleavage sites within a single epitope. Other studies have tested the predictions of internal cleavage sites on its contribution to epitope identification when combined with predictions of major histocompatibility complex (MHC) class I binding; however, it was found that none of the internal sites could improve the ability to identify epitopes (51). Still, it is reasonable that the virus uses this mechanism to escape. For example, in a previous study the binding motifs of HLA-B57 usually carry Phe, Trp, Ile, or Val as its C-terminal residue (4). However, NetChop 3.0 predicts an internal cleavage site on P10 Met as the C terminus of the HLA-B57-restricted epitope KAF11 in p24. This will impair the many interactions between the P11 Phe (the wild-type C terminus) and the contact residues of HLA-B*5703 (81).

MHC affinity prediction algorithms were difficult to integrate into this study because most have limited numbers of HLA class I alleles. In addition, the alleles available in most algorithms were mainly non-African. Furthermore, many dealt exclusively with 9-mers and a few with 10-mers, but HLA-restricted epitopes in *gag* vary from 8-mers to 12-mers. Likewise, the TAP affinity prediction algorithm was also designed for 9-mers; nevertheless, it was applied to peptides with more than nine residues in this study. As mentioned in Materials and Methods, the algorithm was applied in a previous study on peptides with lengths between 10 and 18 amino acids (74). Peters et al. found that the correlation between predicted and measured affinity values for the 10- to 18-mers was lower than for the 9-mers, but it was still significant. Another interesting study also integrated predictions of C-terminal proteasomal cleavage and TAP transport efficiency as well as MHC class I binding affinity to identify CTL epitopes (51); however, their predictions only generated 9-mers, and thus it was difficult to compare their findings with ours.

Negative correlations with HLA alleles and positive selection were frequently observed in our analysis. These negative associations are thought to be a result of successful transmission of positively selected escape mutations to the point of fixation (52). An escape variant that becomes abundant in the population will be lost as a potential target for the immune system. For example, HLA-Cw*0701 is negatively correlated to the I190V mutation in p24 clade A1 sequences. Ile-190 is suspected to be an escape mutation when the epitope is driven by HLA-Cw*0701, a common allele in this East African population. This V190I mutation may have been transmitted and subsequently accumulated within the population, which led to a replacement of Val by Ile as the consensus sequence (52). The increase in CD4 count associated with this replacement may explain one of the reasons why HLA-Cw*0701 is associated with women with reduced susceptibility to HIV-1 infection (20). Previous studies have shown that ultimately not all escape mutations accumulate in a population. Escape mutations that result in a fitness cost usually revert back to the wild type when transmitted to individuals lacking the particular HLA allele that exerted the selection (23, 53).

The p24 capsid is a highly conserved protein and seems to be less tolerant of mutations in comparison to the other residues of Gag. Changes in critical regions may abolish important functions and therefore be detrimental to the virus (56). This study investigated the cyclophilin A binding loop in p24 (residues 217 to 225) (24) and identified a positively selected amino acid substitution at residue 223 from Ile to Val (n = 89) in clade D-infected individuals. However, this amino acid change did not correlate with any change in mean CD4 count (data not shown) and is restricted by HLA-B7 (44). Another amino acid, Asn, in the same residue is also very common (n = 56) but is defined as a neutral drift by QUASI. This mutation is correlated with HLA-B35 (P = 0.004) (data not shown), and it is correlated with a higher mean CD4 count (280.56 compared to 508.62 CD4 cells/ml [$P = 7.31 \times 10^{-4}$]) (Table 7). Because it resides in a region critical to cyclophilin A binding (which is thought to be important in virion disassembly), an Asn mutation at this site may hinder the hydrophobic and van der Waals interactions between these proteins and consequently reduce incorporation of cypA. This could have a negative effect on viral fitness (24).

p6 *gag* is critical in incorporating the accessory protein Vpr into the mature virion, and several studies have attempted to establish the region of the protein that is important in Vpr binding (13, 37). The $(LXX)_4$ region has been implicated in Vpr binding (46); however, mutational analysis indicates that the $_{462}$FRFG region is most important (89). The A1 clade consensus of this study population at this region is $_{462}$FGMG, a sequence that differs notably from the aforementioned $_{462}$FRFG (a sequence which did not occur in A1 clade sequences from this study). Interestingly, the mutation G465R was strongly associated with higher CD4 counts (502.74 cells/ml compared to the consensus 355.85 cells/ml; $P = 4.00 \times 10^{-3}$), indicating that the glycine at position 465 could be functionally important. Similarly, the D clade viruses (the consensus is $_{463}$FGFG) that contain the G464R mutation are associated with a mean CD4 count of 420.82 cells/ml, which is much higher than the CD4 counts for the consensus average of 292.82 cell/ml, although this was not statistically significant

($P = 0.149$; $n = 98$; the variances between the two were unequal [$P = 0.38$], and this rendered the means comparison insignificant). The interaction between Vpr and p6 has not been clearly established, nor has that between Vpr and the nuclear pore complex. There are suggestions that Vpr preferentially interacts with FG repeats in nuclear pore proteins (19), but there has not been a clear consensus reached on this topic. This study suggests the importance for the Gly residues at these locations; however, further functional analysis would need to be conducted to establish its importance in Vpr interaction.

In conclusion, a comprehensive understanding of the relationship between host-restricted selection and immune escape will greatly help in designing a vaccine. Ultimately, an efficient approach to understanding viral evolution would require large-scale sequence analysis. The method outlined in this study is one such approach. The method of identifying positively selected amino acids and studying their relationships with the host (in the context of CTL pressure and their effects on disease progression) plays no small role in increasing our means of identifying potential CTL epitopes and viral adaptation. With this knowledge, detailed mechanisms of immune escape can be inferred and used to identify where the virus may be vulnerable. While this method is a sound and productive informatic approach to studying suspected escape mutations, the predictive algorithms ultimately need to be confirmed with further in vitro functional studies.

ACKNOWLEDGMENTS

This work was supported by the NIH (RO1 A1 49383) and CIHR (HOP-43135). F. A. Plummer is a Tier I Canada Research Chair.

We thank members of the Plummer lab, Philip Lacap, Janis Huntington, Thomas Bielawny, and Will Turk, for their technical assistance and the DNA Core Facility (National Microbiology Lab, Winnipeg, Canada) for the use of the ABI 3730 Genetic Analyzer. We thank the dedicated nurses and staff working in the Pumwani Sex Worker Cohort, Jane Njoki, Jane Kamene, Elizabeth Bwibo, and Edith Amatiwa. We especially thank the women of the Pumwani Sex Worker Cohort for their participation and support. We thank Gary Van Domselaar, Lyle McKinnon, Shehzad Iqbal, T. Blake Ball, Jeffery Tuff, and Allison Land, and Ben Liang for their editorial assistance.

REFERENCES

1. **Addo, M. M., X. G. Yu, A. Rathod, D. Cohen, R. L. Eldridge, D. Strick, M. N. Johnston, C. Corcoran, A. G. Wurcel, C. A. Fitzpatrick, M. E. Feeney, W. R. Rodriguez, N. Basgoz, R. Draenert, D. R. Stone, C. Brander, P. J. Goulder, E. S. Rosenberg, M. Altfeld, and B. D. Walker.** 2003. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. J. Virol. **77:**2081–2092.
2. **Allen, T. M., M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, K. M. O'sullivan, I. Desouza, M. E. Feeney, R. L. Eldridge, E. L. Maier, D. E. Kaufmann, M. P. Lahaie, L. Reyor, G. Tanzi, M. N. Johnston, C. Brander, R. Draenert, J. K. Rockstroh, H. Jessen, E. S. Rosenberg, S. A. Mallal, and B. D. Walker.** 2005. Selective escape from $CD8^+$ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J. Virol. **79:**13239–13249.
3. **Bansal, A., E. Gough, S. Sabbaj, D. Ritter, K. Yusim, G. Sfakianos, G. Aldrovandi, R. A. Kaslow, C. M. Wilson, M. J. Mulligan, J. M. Kilby, and P. A. Goepfert.** 2005. CD8 T-cell responses in early HIV-1 infection are skewed towards high entropy peptides. AIDS **19:**241–250.
4. **Barber, L. D., L. Percival, K. L. Arnett, J. E. Gumperz, L. Chen, and P. Parham.** 1997. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. J. Immunol. **158:**1660–1669.
5. **Barugahare, B., C. Baker, O. K'Aluoch, R. Donovan, M. Elrefaei, M. Eggena, N. Jones, S. Mutalya, C. Kityo, P. Mugyenyi, and H. Cao.** 2005. Human immunodeficiency virus-specific responses in adult Ugandans: patterns of cross-clade recognition. J. Virol. **79:**4132–4139.
6. **Boutwell, C. L., and M. Essex.** 2007. Identification of HLA class I-associated amino acid polymorphisms in the HIV-1C proteome. AIDS Res. Hum. Retrovir. **23:**165–174.
7. **Cao, H., P. Kanki, J. L. Sankale, A. Dieng-Sarr, G. P. Mazzara, S. A. Kalams, B. Korber, S. Mboup, and B. D. Walker.** 1997. Cytotoxic T-lymphocyte cross-reactivity among different human immunodeficiency virus type 1 clades: implications for vaccine development. J. Virol. **71:**8615–8623.
8. **Craiu, A., T. Akopian, A. Goldberg, and K. L. Rock.** 1997. Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. Proc. Natl. Acad. Sci. USA **94:**10850–10855.
9. **Currier, J. R., W. E. Dowling, K. M. Wasunna, U. Alam, C. J. Mason, M. L. Robb, J. K. Carr, F. E. McCutchan, D. L. Birx, and J. H. Cox.** 2003. Detection of high frequencies of HIV-1 cross-subtype reactive CD8 T lymphocytes in the peripheral blood of HIV-1-infected Kenyans. AIDS **17:**2149–2157.
10. **Dawson, L., and X. F. Yu.** 1998. The role of nucleocapsid of HIV-1 in virus assembly. Virology **251:**141–157.
11. **Day, C. L., A. K. Shea, M. A. Altfeld, D. P. Olson, S. P. Buchbinder, F. M. Hecht, E. S. Rosenberg, B. D. Walker, and S. A. Kalams.** 2001. Relative dominance of epitope-specific cytotoxic T-lymphocyte responses in human immunodeficiency virus type 1-infected persons with shared HLA alleles. J. Virol. **75:**6279–6291.
12. **De Groot, A. S., B. Jesdale, W. Martin, C. Saint Aubin, H. Sbai, A. Bosma, J. Lieberman, G. Skowron, F. Mansourati, and K. H. Mayer.** 2003. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine **21:**4486–4504.
13. **Demirov, D. G., A. Ono, J. M. Orenstein, and E. O. Freed.** 2002. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. Proc. Natl. Acad. Sci. USA **99:**955–960.
14. **Dorrell, L., T. Dong, G. S. Ogg, S. Lister, S. McAdam, T. Rostron, C. Conlon, A. J. McMichael, and S. L. Rowland-Jones.** 1999. Distinct recognition of non-clade B human immunodeficiency virus type 1 epitopes by cytotoxic T lymphocytes generated from donors infected in Africa. J. Virol. **73:**1708–1714.
15. **Draenert, R., S. Le Gall, K. J. Pfafferott, 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.
16. **Druillennec, S., A. Caneparo, H. de Rocquigny, and B. P. Roques.** 1999. Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1. J. Biol. Chem. **274:**11283–11288.
17. **El Galta, R., L. Hsu, and J. J. Houwing-Duistermaat.** 2005. Methods to test for association between a disease and a multi-allelic marker applied to a candidate region. BMC Genet. **6**(Suppl. 1)**:**S101.
18. **Evans, D. T., D. H. O'Connor, P. Jing, J. L. Dzuris, J. Sidney, J. da Silva, T. M. Allen, H. Horton, J. E. Venham, R. A. Rudersdorf, T. Vogel, C. D. Pauza, R. E. Bontrop, R. DeMars, A. Sette, A. L. Hughes, and D. I. Watkins.** 1999. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. Nat. Med. **5:**1270–1276.
19. **Fouchier, R. A., B. E. Meyer, J. H. Simon, U. Fischer, A. V. Albright, F. Gonzalez-Scarano, and M. H. Malim.** 1998. Interaction of the human immunodeficiency virus type 1 Vpr protein with the nuclear pore complex. J. Virol. **72:**6004–6013.
20. **Fowke, K. R., N. J. Nagelkerke, J. Kimani, J. N. Simonsen, A. O. Anzala, J. J. Bwayo, K. S. MacDonald, E. N. Ngugi, and F. A. Plummer.** 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet **348:**1347–1351.
21. **Frahm, N., and C. Brander.** 2005. Optimal CTL epitope identification in HIV clade B and non-clade B infection. HIV Mol. Immunol. **I-A:**3–20.
22. **Freed, E. O.** 1998. HIV-1 Gag proteins: diverse functions in the virus life cycle. Virology **251:**1–15.
23. **Friedrich, T. C., E. J. Dodds, L. J. Yant, L. Vojnov, R. Rudersdorf, C. Cullen, D. T. Evans, R. C. Desrosiers, B. R. Mothe, J. Sidney, A. Sette, K. Kunstman, S. Wolinsky, M. Piatak, J. Lifson, A. L. Hughes, N. Wilson, D. H. O'Connor, and D. I. Watkins.** 2004. Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nat. Med. **10:**275–281.
24. **Gamble, T. R., F. F. Vajdos, S. Yoo, D. K. Worthylake, M. Houseweart, W. I. Sundquist, and C. P. Hill.** 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell **87:**1285–1294.
25. **Ganser, B. K., S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist.** 1999. Assembly and analysis of conical models for the HIV-1 core. Science **283:**80–83.
26. **Geels, M. J., S. A. Dubey, K. Anderson, E. Baan, M. Bakker, G. Pollakis, W. A. Paxton, J. W. Shiver, and J. Goudsmit.** 2005. Broad cross-clade T-cell responses to gag in individuals infected with human immunodeficiency virus type 1 non-B clades (A to G): importance of HLA anchor residue conservation. J. Virol. **79:**11247–11258.
27. **Gillespie, G. M., R. Kaul, T. Dong, H. B. Yang, T. Rostron, J. J. Bwayo, P. Kiama, T. Peto, F. A. Plummer, A. J. McMichael, and S. L. Rowland-Jones.**

2002. Cross-reactive cytotoxic T lymphocytes against a HIV-1 p24 epitope in slow progressors with B*57. AIDS **16:**961–972.
28. **Goulder, P. J., M. M. Addo, M. A. Altfeld, E. S. Rosenberg, Y. Tang, U. Govender, N. Mngqundaniso, K. Annamalai, T. U. Vogel, M. Hammond, M. Bunce, H. M. Coovadia, and B. D. Walker.** 2001. Rapid definition of five novel HLA-A*3002-restricted human immunodeficiency virus-specific cytotoxic T-lymphocyte epitopes by Elispot and intracellular cytokine staining assays. J. Virol. **75:**1339–1347.
29. **Goulder, P. J., C. Brander, K. Annamalai, N. Mngqundaniso, U. Govender, Y. Tang, S. He, K. E. Hartman, C. A. O'Callaghan, G. S. Ogg, M. A. Altfeld, E. S. Rosenberg, H. Cao, S. A. Kalams, M. Hammond, M. Bunce, S. I. Pelton, S. A. Burchett, K. McIntosh, H. M. Coovadia, and B. D. Walker.** 2000. Differential narrow focusing of immunodominant human immunodeficiency virus Gag-specific cytotoxic T-lymphocyte responses in infected African and Caucasoid adults and children. J. Virol. **74:**5679–5690.
30. **Goulder, P. J., M. Bunce, P. Krausa, K. McIntyre, S. Crowley, B. Morgan, A. Edwards, P. Giangrande, R. E. Phillips, and A. J. McMichael.** 1996. Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. AIDS Res. Hum. Retrovir. **12:**1691–1698.
31. **Goulder, P. J., A. K. Sewell, D. G. Lalloo, D. A. Price, J. A. Whelan, J. Evans, G. P. Taylor, G. Luzzi, P. Giangrande, R. E. Phillips, and A. J. McMichael.** 1997. Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation. J. Exp. Med. **185:**1423–1433.
32. **Goulder, P. J., Y. Tang, S. I. Pelton, and B. D. Walker.** 2000. HLA-B57-restricted cytotoxic T-lymphocyte activity in a single infected subject toward two optimal epitopes, one of which is entirely contained within the other. J. Virol. **74:**5291–5299.
33. **Goulder, P. J., and B. D. Walker.** 1999. The great escape: AIDS viruses and immune control. Nat. Med. **5:**1233–1235.
34. **Gupta, K., D. Ott, T. J. Hope, R. F. Siliciano, and J. D. Boeke.** 2000. A human nuclear shuttling protein that interacts with human immunodeficiency virus type 1 matrix is packaged into virions. J. Virol. **74:**11811–11824.
35. **Henikoff, S., and J. G. Henikoff.** 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89:**10915–10919.
36. **Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen, and W. I. Sundquist.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc. Natl. Acad. Sci. USA **93:**3099.
37. **Holguin, A., A. Alvarez, and V. Soriano.** 2006. Variability in the $P6^{gag}$ domains of HIV-1 involved in viral budding. AIDS **20:**624–627.
38. **Ikeda-Moore, Y., H. Tomiyama, M. Ibe, S. Oka, K. Miwa, Y. Kaneko, and M. Takiguchi.** 1998. Identification of a novel HLA-A24-restricted cytotoxic T-lymphocyte epitope derived from HIV-1 Gag protein. AIDS **12:**2073–2074.
39. **Jassoy, C., R. P. Johnson, B. A. Navia, J. Worth, and B. D. Walker.** 1992. Detection of a vigorous HIV-1-specific cytotoxic T lymphocyte response in cerebrospinal fluid from infected persons with AIDS dementia complex. J. Immunol. **149:**3113–3119.
40. **Johnson, R. P., A. Trocha, L. Yang, G. P. Mazzara, D. L. Panicali, T. M. Buchanan, and B. D. Walker.** 1991. HIV-1 gag-specific cytotoxic T lymphocytes recognize multiple highly conserved epitopes. Fine specificity of the gag-specific response defined by using unstimulated peripheral blood mononuclear cells and cloned effector cells. J. Immunol. **147:**1512–1521.
41. **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.
42. **Kelleher, A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Olson, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, A. J. McMichael, S. Rowland-Jones, and R. E. Phillips.** 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. **193:**375–386.
43. **Kesmir, C., A. K. Nussbaum, H. Schild, V. Detours, and S. Brunak.** 2002. Prediction of proteasome cleavage motifs by neural networks. Protein Eng. **15:**287–296.
44. **Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, 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.
45. **Kiepiela, P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, B. D. Walker, and P. Goulder.** 2007. $CD8^+$ T-cell responses to different HIV proteins have discordant associations with viral load. Nat. Med. **13:**46–53.
46. **Kondo, E., and H. G. Gottlinger.** 1996. A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. J. Virol. **70:**159–164.
47. **Korber, B. T., 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.
48. **Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya.** 2000. Timing the ancestor of the HIV-1 pandemic strains. Science **288:**1789–1796.
49. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
50. **Lange, K.** 2002. Mathematical and statistical methods for genetic analysis, p. 59–79. Springer-Verlag, New York, NY.
51. **Larsen, M. V., C. Lundegaard, K. Lamberth, S. Buus, S. Brunak, O. Lund, and M. Nielsen.** 2005. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur. J. Immunol. **35:**2295–2303.
52. **Leslie, A., D. Kavanagh, I. Honeyborne, K. Pfafferott, C. Edwards, T. Pillay, L. Hilton, C. Thobakgale, D. Ramduth, R. Draenert, S. Le Gall, G. Luzzi, A. Edwards, C. Brander, A. K. Sewell, S. Moore, J. Mullins, C. Moore, S. Mallal, N. Bhardwaj, K. Yusim, R. Phillips, P. Klenerman, B. Korber, P. Kiepiela, B. Walker, and P. Goulder.** 2005. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. J. Exp. Med. **201:**891–902.
53. **Leslie, A. J., K. J. Pfafferott, P. Chetty, R. Draenert, M. M. Addo, M. Feeney, Y. Tang, E. C. Holmes, T. Allen, J. G. Prado, M. Altfeld, C. Brander, C. Dixon, D. Ramduth, P. Jeena, S. A. Thomas, A. St John, T. A. Roach, B. Kupfer, G. Luzzi, A. Edwards, G. Taylor, H. Lyall, G. Tudor-Williams, V. Novelli, J. Martinez-Picado, P. Kiepiela, B. D. Walker, and P. J. Goulder.** 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat. Med. **10:**282–289.
54. **Lieberman, J., J. A. Fabry, D. M. Fong, and G. R. Parkerson 3rd.** 1997. Recognition of a small number of diverse epitopes dominates the cytotoxic T lymphocytes response to HIV type 1 in an infected individual. AIDS Res. Hum. Retrovir. **13:**383–392.
55. **Luo, M., J. Blanchard, Y. Pan, K. Brunham, and R. C. Brunham.** 1999. High-resolution sequence typing of HLA-DQA1 and -DQB1 exon 2 DNA with taxonomy-based sequence analysis (TBSA) allele assignment. Tissue Antigens **54:**69–82.
56. **Martinez-Picado, J., J. G. Prado, E. E. Fry, K. Pfafferott, A. Leslie, S. Chetty, C. Thobakgale, I. Honeyborne, H. Crawford, P. Matthews, T. Pillay, C. Rousseau, J. I. Mullins, C. Brander, B. D. Walker, D. I. Stuart, P. Kiepiela, and P. Goulder.** 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J. Virol. **80:**3617–3623.
57. **Masemola, A. M., T. N. Mashishi, G. Khoury, H. Bredell, M. Paximadis, T. Mathebula, D. Barkhan, A. Puren, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. McIntyre, S. A. Karim, H. W. Sheppard, C. M. Gray, et al.** 2004. Novel and promiscuous CTL epitopes in conserved regions of Gag targeted by individuals with early subtype C HIV type 1 infection from southern Africa. J. Immunol. **173:**4607–4617.
58. **McKinney, D. M., R. Skvoretz, B. D. Livingston, C. C. Wilson, M. Anders, R. W. Chesnut, A. Sette, M. Essex, V. Novitsky, and M. J. Newman.** 2004. Recognition of variant HIV-1 epitopes from diverse viral subtypes by vaccine-induced CTL. J. Immunol. **173:**1941–1950.
59. **McMichael, A., and P. Klenerman.** 2002. HIV/AIDS. HLA leaves its footprints on HIV. Science **296:**1410–1411.
60. **McMichael, A. J., and S. L. Rowland-Jones.** 2001. Cellular immune responses to HIV. Nature **410:**980–987.
61. **Migueles, S. A., A. C. Laborico, H. Imamichi, W. L. Shupert, C. Royce, M. McLaughlin, L. Ehler, J. Metcalf, S. Liu, C. W. Hallahan, and M. Connors.** 2003. The differential ability of HLA $B*5701^+$ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. J. Virol. **77:**6889–6898.
62. **Milicic, A., C. T. Edwards, S. Hue, J. Fox, H. Brown, T. Pillay, J. W. Drijfhout, J. N. Weber, E. C. Holmes, S. J. Fidler, H. T. Zhang, and R. E. Phillips.** 2005. Sexual transmission of single human immunodeficiency virus type 1 virions encoding highly polymorphic multisite cytotoxic T-lymphocyte escape variants. J. Virol. **79:**13953–13962.
63. **Mo, X. Y., P. Cascio, K. Lemerise, A. L. Goldberg, and K. Rock.** 1999. Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. J. Immunol. **163:**5851–5859.
64. **Morikawa, Y., W. H. Zhang, D. J. Hockley, M. V. Nermut, and I. M. Jones.** 1998. Detection of a trimeric human immunodeficiency virus type 1 Gag intermediate is dependent on sequences in the matrix protein, p17. J. Virol. **72:**7659–7663.
65. **Musey, L., Y. Ding, J. Cao, J. Lee, C. Galloway, A. Yuen, K. R. Jerome, and M. J. McElrath.** 2003. Ontogeny and specificities of mucosal and blood

human immunodeficiency virus type 1-specific CD8$^+$ cytotoxic T lymphocytes. J. Virol. **77:**291–300.
66. **Musey, L., Y. Hu, L. Eckert, M. Christensen, T. Karchmer, and M. J. McElrath.** 1997. HIV-1 induces cytotoxic T lymphocytes in the cervix of infected women. J. Exp. Med. **185:**293–303.
67. **Nielsen, M., C. Lundegaard, O. Lund, and C. Kesmir.** 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics **57:**33–41.
68. **Novitsky, V., H. Cao, N. Rybak, P. Gilbert, M. F. McLane, S. Gaolekwe, T. Peter, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee, and M. Essex.** 2002. Magnitude and frequency of cytotoxic T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. J. Virol. **76:**10155–10168.
69. **Novitsky, V., N. Rybak, M. F. McLane, P. Gilbert, P. Chigwedere, I. Klein, S. Gaolekwe, S. Y. Chang, T. Peter, I. Thior, T. Ndung'u, F. Vannberg, B. T. Foley, R. Marlink, T. H. Lee, and M. Essex.** 2001. Identification of human immunodeficiency virus type 1 subtype C Gag-, Tat-, Rev-, and Nef-specific elispot-based cytotoxic T-lymphocyte responses for AIDS vaccine design. J. Virol. **75:**9210–9228.
70. **Ono, A., J. M. Orenstein, and E. O. Freed.** 2000. Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. J. Virol. **74:**2855–2866.
71. **Overbaugh, J., J. Kreiss, M. Poss, P. Lewis, S. Mostad, G. John, R. Nduati, D. Mbori-Ngacha, H. Martin, Jr., B. Richardson, S. Jackson, J. Neilson, E. M. Long, D. Panteleeff, M. Welch, J. Rakwar, D. Jackson, B. Chohan, L. Lavreys, K. Mandaliya, J. Ndinya-Achola, and J. Bwayo.** 1999. Studies of human immunodeficiency virus type 1 mucosal viral shedding and transmission in Kenya. J. Infect. Dis. **179**(Suppl. 3)**:**S401–S404.
72. **Parker, K. C., M. A. Bednarek, L. K. Hull, U. Utz, B. Cunningham, H. J. Zweerink, W. E. Biddison, and J. E. Coligan.** 1992. Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. J. Immunol. **149:**3580–3587.
73. **Paz, P., N. Brouwenstijn, R. Perry, and N. Shastri.** 1999. Discrete proteolytic intermediates in the MHC class I antigen processing pathway and MHC I-dependent peptide trimming in the ER. Immunity **11:**241–251.
74. **Peters, B., S. Bulik, R. Tampe, P. M. Van Endert, and H. G. Holzhutter.** 2003. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J. Immunol. **171:**1741–1749.
75. **Pond, S. L., and S. D. Frost.** 2005. DataMonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21:**2531–2533.
76. **Schmalzbauer, E., B. Strack, J. Dannull, S. Guehmann, and K. Moelling.** 1996. Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. J. Virol. **70:**771–777.
77. **Serwold, T., S. Gaw, and N. Shastri.** 2001. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. Nat. Immunol. **2:**644–651.
78. **Sham, P. C., and D. Curtis.** 1995. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann. Hum. Genet. **59:**97–105.
79. **Shannon, C. E.** 1948. A mathematical theory of communication. Bell Telephone System Technical J. **27:**379–423, 623–656.
80. **Stewart, J. J., P. Watts, and S. Litwin.** 2001. An algorithm for mapping positively selected members of quasispecies-type viruses. BMC Bioinformatics **2:**1.
81. **Stewart-Jones, G. B., G. Gillespie, I. M. Overton, R. Kaul, P. Roche, A. J. McMichael, S. Rowland-Jones, and E. Y. Jones.** 2005. Structures of three HIV-1 HLA-B*5703-peptide complexes and identification of related HLAs potentially associated with long-term nonprogression. J. Immunol. **175:**2459–2468.
82. **Stoltze, L., T. P. Dick, M. Deeg, B. Pommerl, H. G. Rammensee, and H. Schild.** 1998. Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. Eur. J. Immunol. **28:**4029–4036.
83. **Vajdos, F. F., S. Yoo, M. Houseweart, W. I. Sundquist, and C. P. Hill.** 1997. Crystal structure of cyclophilin A complexed with a binding site peptide from the HIV-1 capsid protein. Protein Sci. **6:**2297–2307.
84. **von Schwedler, U. K., M. Stuchell, B. Muller, D. M. Ward, H. Y. Chung, E. Morita, H. E. Wang, T. Davis, G. P. He, D. M. Cimbora, A. Scott, H. G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist.** 2003. The protein network of HIV budding. Cell **114:**701–713.
85. **Wu, Z., J. Alexandratos, B. Ericksen, J. Lubkowski, R. C. Gallo, and W. Lu.** 2004. Total chemical synthesis of N-myristoylated HIV-1 matrix protein p17: structural and mechanistic implications of p17 myristoylation. Proc. Natl. Acad. Sci. USA **101:**11587–11592.
86. **Yang, O. O., P. T. Sarkis, A. Ali, J. D. Harlow, C. Brander, S. A. Kalams, and B. D. Walker.** 2003. Determinant of HIV-1 mutational escape from cytotoxic T lymphocytes. J. Exp. Med. **197:**1365–1375.
87. **Yu, X. G., M. M. Addo, E. S. Rosenberg, W. R. Rodriguez, P. K. Lee, C. A. Fitzpatrick, M. N. Johnston, D. Strick, P. J. Goulder, B. D. Walker, and M. Altfeld.** 2002. Consistent patterns in the development and immunodominance of human immunodeficiency virus type 1 (HIV-1)-specific CD8$^+$ T-cell responses following acute HIV-1 infection. J. Virol. **76:**8690–8701.
88. **Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber.** 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J. Virol. **76:**8757–8768.
89. **Zhu, H., H. Jian, and L. J. Zhao.** 2004. Identification of the 15FRFG domain in HIV-1 Gag p6 essential for Vpr packaging into the virion. Retrovirology **1:**26.