JOURNAL OF VIROLOGY, Sept. 2008, p. 8548–8559 Vol. 82, No. 17
0022-538X/08/$08.00+0 doi:10.1128/JVI.00580-08

Central Role of Reverting Mutations in HLA Associations with Human Immunodeficiency Virus Set Point▽

Philippa C. Matthews,[1]†* Andrew Prendergast,[1]† Alasdair Leslie,[1] Hayley Crawford,[1] Rebecca Payne,[1] Christine Rousseau,[2] Morgane Rolland,[2] Isobella Honeyborne,[1] Jonathan Carlson,[3] Carl Kadie,[3] Christian Brander,[4] Karen Bishop,[5] Nonkululeko Mlotshwa,[5] James I. Mullins,[2] Hoosen Coovadia,[5] Thumbi Ndung'u,[5] Bruce D. Walker,[4,6] David Heckerman,[3] and Philip J. R. Goulder[1,4,5]*

Department of Paediatrics, Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, South Parks Rd., Oxford OX1 3SY, United Kingdom[1]; Department of Microbiology, University of Washington School of Medicine, Seattle, Washington 98195-8070[2]; Microsoft Research, One Microsoft Way, Redmond, Washington 98052[3]; Partners AIDS Research Center, Massachusetts General Hospital, 13th St., Bldg. 149, Charlestown, Boston, Massachusetts 02129[4]; HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa[5]; and Howard Hughes Medical Institute, Chevy Chase, Maryland[6]

Received 14 March 2008/Accepted 10 June 2008

Much uncertainty still exists over what T-cell responses need to be induced by an effective human immunodeficiency virus (HIV) vaccine. Previous studies have hypothesized that the effective CD8$^+$ T-cell responses are those driving the selection of escape mutations that reduce viral fitness and therefore revert posttransmission. In this study, we adopted a novel approach to define better the role of reverting escape mutations in immune control of HIV infection. This analysis of sequences from 710 study subjects with chronic C-clade HIV type 1 infection demonstrates the importance of mutations that impose a fitness cost in the control of viremia. Consistent with previous studies, the viral set points associated with each HLA-B allele are strongly correlated with the number of Gag-specific polymorphisms associated with the relevant HLA-B allele ($r = -0.56$, $P = 0.0034$). The viral set points associated with each HLA-C allele were also strongly correlated with the number of Pol-specific polymorphisms associated with the relevant HLA-C allele ($r = -0.67$, $P = 0.0047$). However, critically, both these correlations were dependent solely on the polymorphisms identified as reverting. Therefore, despite the inevitable evolution of viral escape, viremia can be controlled through the selection of mutations that are detrimental to viral fitness. The significance of these results is in highlighting the rationale for an HIV vaccine that can induce these broad responses.

The recent failure of the first T-cell-based human immunodeficiency virus (HIV) vaccine (34, 43) has emphasized the urgent need to refocus on the question of what T-cell responses need to be induced by an HIV vaccine. One of the strongest clues to immune control of HIV comes from the consistent associations observed between the expression of particular HLA class I alleles, such as HLA-B*57 or -B*27, and a low viral set point and between other alleles, such as HLA-B*35 and -B*5802, and a high viral set point (13, 18, 19, 32, 36). The HIV-specific CD8$^+$ T-cell responses generated in infected individuals who have HLA-B*57 or -B*27 are dominated, in both acute and chronic infections, by CD8$^+$ T cells (cytotoxic T lymphocytes [CTL]) that target Gag epitopes (1, 12, 15, 17, 31, 35).

A critical role of Gag-specific responses in the control of viremia has been established in studies of both simian immunodeficiency virus (SIV) and HIV (13, 24, 29, 37, 38, 44), and population-based studies show that, irrespective of HLA type, broad Gag-specific CD8$^+$ T-cell responses are strongly associated with a decreasing viral load (3, 23). Furthermore, the observation that HLA alleles such as B*57 and B*27 select escape mutations within Gag epitopes that impose high fitness costs on the virus, and that therefore are likely to "revert" back to the wild type following transmission to HLA-mismatched recipients (4, 28, 42), suggests a potentially important role for the selection of reverting escape mutations (10, 25).

In addition to HLA-B and broad Gag-specific CD8$^+$ T-cell responses, a genome-wide association study recently indicated that HLA-C-restricted responses may also play an important role in the immune control of HIV (9). Viral kinetic data have shown that not only Gag-specific (40) but also Pol-specific CD8$^+$ T cells are activated by virus-infected cells prior to the de novo synthesis of viral proteins and Nef-mediated major histocompatibility complex (MHC) class I downregulation (41).

* Corresponding author. Mailing address for Philippa Matthews: Department of Paediatrics, Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, South Parks Rd., Oxford OX1 3SY, United Kingdom. Phone: 44 1865 281883. Fax: 44 1865 281890. E-mail: p.matthews@doctors.org.uk. Mailing address for Philip Goulder: Department of Paediatrics, Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, South Parks Rd., Oxford OX1 3SY, United Kingdom. Phone: 44 1865 281884. Fax: 44 1865 281890. E-mail: philip.goulder@paediatrics.ox.ac.uk.

† P.C.M. and A.P. contributed equally to this work.

▽ Published ahead of print on 2 July 2008.

The aim of the present study was therefore to address the roles of HLA-C and Pol-specific responses, and also that of HLA-mediated selection of viral escape mutants that revert posttransmission, in the immune control of HIV. HLA-C-restricted $CD8^+$ T-cell responses have been much understudied compared to HLA-A and HLA-B (16). Furthermore, $CD8^+$ T-cell responses to Pol have been less well studied than Gag and Nef, so it was critically important to employ methodology free of any bias that would result from studying associations that fall only within previously defined epitopes. The approach adopted therefore incorporates no such assumptions.

We employed a refined method that appropriately takes into account factors including multiple tests, viral sequence relatedness, and linkage disequilibrium between different HLA class I molecules and that can also detect identical HIV amino acid polymorphisms associated with distinct HLA class I molecules. In particular, this analysis also uses statistical methodology to identify mutations likely to revert following transmission to an HLA-mismatched recipient. We analyzed viral Gag, Pol, and Nef sequences from 710 study subjects with HIV infection from Durban, South Africa, and $CD8^+$ T-cell responses to all HIV proteins in 681 study subjects from the same cohort.

MATERIALS AND METHODS

Patient cohorts. Seven hundred ten treatment-naïve adults with C-clade HIV infection were recruited from Durban, South Africa, following voluntary counseling and testing in either antenatal or outpatient clinics as previously described (23). Viral loads and CD4 counts were available for 681 of these study subjects. Viral load was measured with the Roche Amplicor version 1.5 assay, and the CD4 count was measured by flow cytometry. The median viral load of this cohort was 37,200 HIV RNA copies/ml plasma (interquartile range, 6,485 to 133,500), and the median CD4 count was 387 cells/mm^3 (interquartile range, 257 to 520). High-resolution HLA typing was undertaken with genomic DNA by single-stranded conformation polymorphism PCR (5). In a minority of instances, resolution only to two-digit types was possible. HLA-associated polymorphisms were sought by using both two- and four-digit HLA types. This study was approved by the University of KwaZulu-Natal Review Board, the Oxford Research Ethics Committee, and the MGH Review Board.

Enzyme-linked immunospot (ELISPOT) assays. $CD8^+$ T-cell responses were determined for 681 subjects from this cohort. As previously described, 410 18-mer peptides, spanning the HIV proteome, were synthesized based on the 2001 C-clade consensus sequence (23) and were used in a matrix system of 11 or 12 peptides/pool to screen subjects for HIV-specific T-cell responses by gamma interferon (IFN-γ) ELISPOT assay (22, 23).

Sequencing of proviral DNA and viral RNA. Genomic DNA was extracted from peripheral blood mononuclear cells and amplified by nested PCR as previously described (26). Regions of the virus that were amplified with previously published primers (26) were Gag, Pol, and Nef. The PCR product was purified by polyethylene glycol precipitation and sequenced directly. Sequencing was undertaken with the Big Dye ready reaction termination mix (V3) (Applied Biosystems) as previously described (26, 27). In this way, viral Gag, Pol, and Nef sequences were obtained from 616, 446, and 436 study subjects, respectively. In addition, viral sequencing from plasma RNA was undertaken as previously described (39). RNA was extracted from plasma, reverse transcribed, and amplified by nested PCR; the amplified HIV type 1 genome was then purified, cloned, and sequenced. In this way, viral Gag, Pol, and Nef sequences were obtained from 57, 15, and 7 study subjects. Thus, from the total of 710 study subjects, we generated 673 Gag, 461 Pol, and 443 Nef sequences.

Identification of HLA class I associations with HIV amino acid sequence polymorphisms. An extension of method two described by Bhattacharya et al. (2) was used to identify HLA-I associated HIV polymorphisms. In contrast to earlier studies (33), this approach was refined in order to optimize the detection of true HLA associations with HIV sequence polymorphism (2, 6, 14). First, our approach corrected for the phylogenetic structure of the sequences. Second, we computed P and q values for every association; a q value is used to correct for multiple comparisons by calculation of a false detection rate. Third, we sought to correct for linkage disequilibrium between HLA alleles by inclusion of multiple HLA predictors, allowing for the elimination of spurious associations.

For each protein analyzed, a maximum-likelihood phylogenetic tree was constructed from the corresponding sequences. For every HLA allele, amino acid position, and amino acid at that position, two generative, or directed graphical, models of the observed presence or absence of the amino acid in each sequence were created—one representing the null hypothesis that the observations are generated by the phylogenetic tree alone and the other representing the alternative hypothesis that additional escape or reversion takes place due to HLA pressure in the subjects for which the sequences are observed. The likelihood of the observations was then maximized over the parameters of both models with an expectation maximization algorithm, and a P value was computed with a likelihood ratio test based on those likelihoods. To increase power, the tests were made binary such that the presence or absence of a given HLA allele was correlated with the presence or absence of a given amino acid. In addition, HLA polymorphism pairs were analyzed only when the actual or expected count in every cell of the corresponding two-by-two contingency table was greater than or equal to three. (The number three was chosen by optimizing power on another data set.) For every amino acid at each position, the HLA allele with the strongest association (and its corresponding P value) was added to the list of identified associations. The analysis was then repeated after removing individuals having or possibly having this HLA allele. This procedure was iterated until no HLA allele yielded an association with a P value of less than 0.05. By this approach, the model accounts for the potential overlap of HLA associations at a given residue due to clustering of epitopes.

A q value statistic, estimating the proportion of false positives among the associations identified, was computed for each association by repeating this analysis on null data (generated by permuting the HLA data). The q value statistics were estimated separately for Gag, Pol, and Nef associations by using 1,000,000 null tests per protein. We corrected for multiple comparisons by using both $q < 0.05$ (estimating 5% false positives) and $q < 0.2$ (estimating 20% false positives).

Some sequences contained gaps, and in order to maximize the data, sequence alignments were undertaken by using separate phylogenetic trees for p17 (n = 598), p24 (n = 666), and p15 Gag (n = 433) and for the following fragments within Pol: nucleotides 1 to 969 (n = 394), nucleotides 970 to 2208 (n = 417), and nucleotide 2209 to the end (n = 262). Nef was analyzed as a single fragment (n = 443). Sequences were aligned by Se_Al. The total number of amino acid residues analyzed was 541 for Gag, 1,044 for Pol, and 225 for Nef, including insertions.

This analysis revealed statistical associations between HLA class I expression and HIV sequence polymorphisms, which we used to infer the biological processes of escape and reversion. "Escape" describes amino acid selection away from consensus in the presence of a specific HLA allele, and "reversion" describes amino acid selection toward consensus in the absence of a specific HLA allele. Using methodology previously described (2), this model detects escape as a statistical association between the expression of a particular HLA allele with either (i) a decrease in the number of the consensus amino acid or (ii) an increase in the number of a variant amino acid. Within the category of escape polymorphisms, we also included negative associations, whereby the presence of an HLA allele is associated with consensus. These associations may arise in the context of escape mutations that do not revert, such that the wild type is eventually replaced by the HLA-selected variant and a new population consensus is established (26, 33).

Reversion was detected statistically by association between the absence of a particular HLA allele with either (i) an increase in the number of the consensus amino acid or (ii) a decrease in the number of the variant amino acid. These associations suggest that a mutation selected by a particular HLA type returns to consensus following transmission to an HLA-mismatched host due to a fitness cost imposed by the mutation.

When two alleles are in linkage disequilibrium and only one is actually associated with a polymorphism, the algorithm sometimes assigned the effect to the wrong allele by chance. To correct for this effect, if one of the alleles was associated with a polymorphic residue contained within a described epitope (listed by HIV immunology databases at http://www.hiv.lanl.gov or by Kiepiela et al. [23]) but the other was not, then the first allele was designated as the associated HLA type.

Identification of HLA class I associations with 18-mer peptide recognition. Associations between the recognition of individual 18-mer peptides and the expression of particular HLA class I alleles was sought with Fisher's exact test as previously undertaken (23). For each peptide recognized, a comparison was made of the number of subjects with or without a particular HLA allele who showed recognition of the peptide versus those with or without that HLA allele who did not show recognition of that peptide. All 65 HLA class I alleles ex-

TABLE 1. HLA-A class I frequencies, median viral loads, and numbers of HLA-associated polymorphisms at $q < 0.05$

HLA class I allele or statistical parameter	Phenotypic frequency of allele (%)	Median viral load for allele	No. of HLA-associated polymorphisms at $q < 0.05$		
			Gag	Pol	Nef
A*0101	4.71	22,100	1	0	0
A*0201	6.42	20,000	0	0	0
A*0202	2.14	32,250	0	0	0
A*0205	7.85	24,250	0	0	0
A*0301	7.56	53,650	0	0	0
A*2301	13.55	29,300	0	0	1
A*2402	5.28	38,950	0	0	1
A*26(01)	2.14	35,850	0	0	0
A*2902	1.57	28,150	0	0	0
A*3001	16.12	32,700	0	0	0
A*3002	13.55	55,550	0	1	0
A*3004	2.85	57,650	0	0	0
A*32(01)	1.00	24,100	0	0	0
A*33	3.85	33,800	0	0	0
A*34(02)	4.99	40,900	0	1	0
A*3601	0.86	32,500	0	0	0
A*4301	4.56	38,200	0	0	0
A*6601	5.85	37,750	0	0	0
A*6602	1.14	73,100	0	0	0
A*6801	5.85	72,600	0	0	0
A*6802	16.83	41,000	0	0	0
A*74	9.70	8,180	1	1	1
A*8001	1.14	74,500	0	0	0
Total			2	3	3
Correlations[a]					
Whole data set			$\underline{r = -0.47, P = 0.03}$	$r = 0.00, P = 1.0$	$r = -0.25, P = 0.24$
Reverting sites			$r = -0.35, P = 0.098$	$r = 0.23, P = 0.30$	$r = -0.19, P = 0.40$
Nonreverting sites			$r = -0.29, P = 0.18$	$r = -0.16, P = 0.46$	$r = -0.16, P = 0.46$

[a] Values indicating significant differences are underlined.

pressed at a phenotypic frequency of ≥0.5% were included in the analysis (23 HLA-A, 25 HLA-B, and 17 HLA-C alleles; Tables 1, 2, and 3). To correct for multiple comparisons, the conservative Bonferroni correction was used in order to make this analysis comparable to previous studies (23). Thus, associations of $0.05 > P > 0.00075$ were lost following correction for the 65 alleles analyzed. We also repeated this analysis by computation of q values in place of the Bonferroni correction. A q value of 0.016 (corresponding to a P value of 0.00075 from a univariate model) was used as the cutoff for this analysis.

RESULTS

HLA-associated HIV amino acid polymorphisms in 710 HIV-infected subjects from KwaZulu-Natal, South Africa. In order to take account of the possibility that certain HLA class I molecules are associated with characteristic viral set points as a result of the selection of particular escape mutants in early infection, we set out to determine amino acid polymorphisms in the three most immunogenic proteins, Gag, Pol, and Nef, that are associated with any of the 65 HLA class I molecules expressed at a ≥0.5% phenotypic frequency in this cohort.

By using a novel multivariate analysis correcting for phylogenetic relatedness of sequences (2), for multiple tests, and for linkage disequilibrium between HLA class I alleles and with 5% expected false positives ($q < 0.05$), we identified 84 associations between the expression of particular HLA class I alleles and particular HIV amino acid polymorphisms; the numbers of polymorphisms associated with the individual alleles are summarized in Tables 1 to 3. Overall, 61 of these polymorphisms (73%) were HLA-B associated. Although the majority (58%) of the 205 epitopes targeted to a substantial level in this cohort were also HLA-B restricted (see below), the number of HIV amino acid polymorphisms associated with HLA-B was nonetheless significantly higher than expected ($P = 0.023$ [Fisher's exact test]). Figure 1 summarizes the data for grouped HLA-A, HLA-B, and HLA-C alleles, including analysis with a 5% expected false-positive rate ($q < 0.05$; Fig. 1A and B) and a 20% expected false-positive rate ($q < 0.20$; Fig. 1C and D).

Correlation between HLA-associated polymorphisms and a reducing viral load. We then addressed the question of whether the number of HLA-associated polymorphisms in Gag, Pol, or Nef is related to the median viral load for each of the different HLA alleles. As anticipated, the number of HLA-B-associated polymorphisms in Gag was strongly associated with a reducing viral load ($q < 0.05$; $r = -0.57$, $P = 0.0034$). The proportion of polymorphisms identified within known epitopes restricted by the relevant HLA allele was reduced by increasing the number of false positives included to 20% ($q < 0.2$; Fig. 1D), but the correlation remained unaffected ($r = -0.51$, $P = 0.0096$; Fig. 2B). No such association was identified between the number of HLA-B polymorphisms in Pol or Nef and the median viral load for each allele (data not shown). Unexpectedly, a strong correlation was also observed between the number of polymorphisms within Pol selected by HLA-C alleles and the median viral load for each allele, which was again largely unaffected by the q value chosen ($q < 0.05$, $r =$

TABLE 2. HLA-B class I frequencies, median viral loads, and numbers of HLA-associated polymorphisms at $q < 0.05$

HLA class I allele or statistical parameter	Phenotypic frequency of allele (%)	Median viral load for allele	No. of HLA-associated polymorphisms at $q < 0.05$		
			Gag	Pol	Nef
B*0702	6.70	27,900	1	1	0
B*0705	0.86	86,600	0	0	0
B*0801	8.13	56,900	0	1	0
B*13	3.14	14,650	—[b]	—	—
B*1302	1.57	23,050	0	0	0
B*1401	3.85	21,900	1	0	0
B*1402	1.14	76,200	0	0	0
B*1503	15.26	55,300	0	2	0
B*1510	17.12	51,700	1	1	0
B*1516	1.14	15,017	0	0	0
B*1801	5.56	84,900	0	2	1
B*3501	3.14	40,000	1	1	2
B*3910	4.14	23,850	2	2	0
B*41	0.71	139,000	1	0	0
B*4201	14.98	25,600	0	4	0
B*4202	2.00	109,000	0	0	0
B*4403	9.42	27,600	1	7	3
B*4501	7.85	41,000	0	0	1
B*4901	1.43	157,500	0	0	0
B*51(01)	0.71	135,500	0	0	0
B*5301	3.99	21,700	0	1	1
B*5702	2.28	15,190	3	1	0
B*5703	5.99	7,642	6	0	0
B*5801	12.41	16,700	2	1	1
B*5802	21.26	60,200	0	0	0
B*8101	11.27	12,800	3	2	1
Total			22	26	10
Correlations[a]					
Whole data set			$r = -0.56$, $P = 0.0034$	$r = -0.29$, $P = 0.17$	$r = -0.1$, 5 $P = 0.48$
Reverting sites			$r = -0.62$, $P = 0.0009$	$r = -0.30$, $P = 0.14$	$r = -0.07$, $P = 0.74$
Nonreverting sites			$r = -0.14$, $P = 0.51$	$r = -0.05$, $P = 0.83$	$r = -0.2$, 0 $P = 0.34$

[a] Values indicating significant differences are underlined.
[b] —, HLA-B*13 at two-digit resolution was used only to analyze ELISPOT responses; sequence analysis identified associations at four-digit resolution.

-0.67 and $P = 0.005$; $q < 0.2$, $r = -0.57$ and $P = 0.02$; Fig. 2C and D). Potential correlations between polymorphisms selected by HLA-A alleles and the median viral load for the allele were more difficult to assess, since these were so few at low q values and did not approach significance at higher q values (Table 1).

In order to establish whether the associations observed in Gag were, in part, a consequence of the inclusion of more sequences for Gag than for other proteins, we reduced our pool of Gag sequences, by random selection, to 430 (equivalent to the sequence numbers of the other two proteins) and repeated the lineage-corrected analysis. We found that the significant association between the number of HLA-B responses and the median viral set point was not weakened by this reduction in sequence numbers ($q < 0.05$, $r = -0.63$ and $P = 0.0007$; $q < 0.2$, $r = -0.59$ and $P = 0.002$; data not shown).

The number of polymorphisms that revert following transmission is correlated with a reducing viral load. Previous studies have suggested that the selection of escape mutations that reduce the viral replicative capacity, although allowing escape from the particular $CD8^+$ T-cell specificity, may contribute to subsequent control through the remaining immune responses (4, 8, 12, 21, 27, 42). Therefore, we next examined the impact on the viral load of polymorphisms identified statistically as "reverting" (that is, in the absence of a specific allele, the polymorphism is absent). Reversion suggests that the mutation occurs at a detriment to viral infectivity or replicative capacity, such that a return to the wild type is favored when the HLA selection pressure is removed.

We found a strong correlation between the median viral loads of different HLA-B alleles and the number of reverting HLA-B-associated polymorphisms in Gag (for associations at $q < 0.05$, $r = -0.62$ and $P = 0.0009$; Fig. 3A) but not with nonreverting HLA-B polymorphisms in Gag ($r = -0.14$ and $P = 0.51$; Fig. 3C). This relationship was also confirmed for associations with $q < 0.2$ (Fig. 3B and D). Within Pol, there was a difference between reverting and nonreverting HLA-B-associated polymorphisms (for associations at $q < 0.05$, [i] $r = -0.30$ and $P = 0.14$ and [ii] $r = 0.05$ and $P = 0.83$, respectively) and between reverting and nonreverting HLA-C-restricted polymorphisms ([i] $r = -0.57$ and $P = 0.020$ and [ii] $r = -0.32$ and $P = 0.22$, respectively; Tables 2 and 3). For HLA-A-associated polymorphisms, the numbers are too low to justify extensive analysis but even here there is an indication that reverting mutations are critical to HLA-associated immune control (Table 1). These data support the hypothesis that the escape mutations that reduce viral fitness are those most strongly associated with the control of viremia.

TABLE 3. HLA-C class I frequencies, median viral loads, and numbers of HLA-associated polymorphisms at $q < 0.05$

HLA class I allele or statistical parameter	Phenotypic frequency of allele (%)	Median viral load for allele	No. of HLA-associated polymorphisms at $q < 0.05$		
			Gag	Pol	Nef
Cw*02	18.69	50,800	0	0	0
Cw*0302	2.00	18,800	0	1	0
Cw*0304	0.29	39,700	1	0	0
Cw*04(01)	10.56	30,750	0	2	0
Cw*0501	1.14	80,800	0	0	0
Cw*0602	27.25	49,875	0	0	0
Cw*0701	17.69	33,800	0	0	1
Cw*0702	5.71	31,100	0	0	0
Cw*0704	2.00	152,500	1	0	0
Cw*0801	3.57	52,425	0	0	0
Cw*0802	3.85	44,550	0	0	0
Cw*0804	1.43	ND[a]	0	0	0
Cw*1203	2.85	20,700	0	1	0
Cw*1601	9.56	56,900	0	0	1
Cw*1604	1.28	22,500	1	0	0
Cw*1701	23.25	33,200	0	0	0
Cw*1801	7.70	23,200	0	1	0
Total			3	5	2
Correlations[b]					
Whole data set			$r = 0.087$, $P = 0.75$	$r = -0.67$, $P = 0.0047$	$r = 0.21$, $P = 0.45$
Reverting sites			$r = 0.33$, $P = 0.21$	$r = -0.57$, $P = 0.020$	$r = -0.03$, $P = 0.92$
Nonreverting sites			$r = -0.31$, $P = 0.25$	$r = -0.32$, $P = 0.22$	$r = 0.31$, $P = 0.25$

[a] ND, no data.
[b] Values indicating significant differences are underlined.

We next repeated all of our analyses with the median CD4 count rather than the viral load for each allele. Again, there was a significant correlation between the number of HLA-associated ($q < 0.05$) polymorphisms and the median CD4 count for HLA-B associations in Gag ($r = 0.61$, $P = 0.001$) that was significant for reverting polymorphisms ($r = 0.65$, $P = 0.0005$) but not for nonreverting sites ($r = 0.23$, $P = 0.28$). As for the viral load, there was also a significant association between HLA-C-associated polymorphisms in Pol and the median CD4 count ($r = 0.65$, $P = 0.006$).

The number of HLA-B-restricted Gag epitopes targeted is correlated with a reducing viral load. In order to address the

TABLE 4. Sites of sequence polymorphism in HIV-1 Gag associated with host HLA-B

Protein	HLA	Polymorphism location[a]	*P* value	*q* value	Reversion
p17 Gag	B*41	ATLYCVHEKI**E**VRDTKEALDK	9.41E-05	0.05	
p24 Gag	B*1510 [A*6802]	QNLQGQMVHQ**A**ISPRTLNAWV	3.67E-04	0.05	R
	B*5702	QNLQGQMVHQ**A**ISPRTLNAWV	2.46E-08	0.00	R
	B*5703 [C*0701]	QNLQGQMVHQ**A**ISPRTLNAWV	2.51E-14	0.00	R
	B*5702	NLQGQMVHQA**I**SPRTLNAWVK	7.71E-05	0.01	R
	B*5703	NLQGQMVHQA**I**SPRTLNAWVK	1.62E-08	0.00	R
	B*5703 [A*0202]	NAWVKVIEEK**A**FSPEVIPMFT	1.39E-15	0.00	
	B*5703	WVKVIEEKAF**S**PEVIPMFTAL	2.92E-04	0.04	
	B*81(01) [A*0101]	EVIPMFTALS**E**GATPQDLNTM	4.26E-07	0.00	R
	B*8101	FTALSEGATP**Q**DLNTMLNTVG	3.59E-05	0.01	R
	B*8101	SEGATPQDLN**T**MLNTVGGHQA	5.03E-07	0.00	R
	B*5702	RGSDIAGTTS**T**LQEQIAWMTS	4.27E-13	0.00	R
	B*5703	RGSDIAGTTS**T**LQEQIAWMTS	4.04E-23	0.00	R
	B*5801	RGSDIAGTTS**T**LQEQIAWMTS	1.88E-28	0.00	R
	B*5703	AGTTSTLQEQ**I**AWMTSNPPIP	1.73E-05	0.00	R
	B*35(01)	MTSNPPIPVG**D**IYKRWIILGL	1.09E-07	0.00	
	B*1401	PFRDYVDRFF**K**TLRAEQATQD	2.54E-11	0.00	R
	B*5801	FFKTLRAEQA**T**QDVKNWMTDT	3.73E-07	0.00	R
	B*4403	KTLRAEQATQ**D**VKNWMTDTLL	3.03E-09	0.00	
	B*3910	LVQNANPDCK**T**ILRALGPGAT	1.36E-04	0.01	
	B*0702	MTACQGVGGP**G**HKARVL	6.72E-10	0.03	
p15 Gag	B*3910	QNRPEPRPEP**T**APPAENFRES	2.03E-04	0.00	R

[a] Boldface indicates site of polymorphism (consensus amino acid listed). Underlining indicates sites of known HLA epitopes restricted by the selecting allele.

TABLE 5. Sites of sequence polymorphism in HIV-1 Pol associated with host HLA-B

Protein	HLA	Polymorphism location[a]	*P* value	*q* value	Reversion
Protease	B*5301	LAFPQQGGEA**R**EFPSEQTRAN	6.62E-06	0.00	
	B*4403	DTGADDTVLE**E**INLPGKWKPK	3.13E-08	0.00	
	B*44(03)	GADDTVLEEI**N**LPGKWKPKMI	1.55E-04	0.04	
RT	B*8101	LGCTLNFPIS**P**IETVPVKLKP	7.40E-10	0.00	R
	B*8101 [Cw*18]	CTLNFPISPI**E**TVPVKLKPGM	1.53E-05	0.00	R
	B*35(01)	GDAYFSVPLD**E**GFRKYTAFTI	5.39E-07	0.00	
	B*1801	TAFTIPSINN**E**TPGIRYQYNV	4.86E-05	0.05	
	B*0702	GWKGSPAIFQ**S**SMTKILEPFR	1.93E-06	0.00	
	B*4403	QHRAKIEELR**E**HLLKWGFTTP	9.00E-07	0.00	R
	B*4201 [Cw*1701]	GKLNWASQIY**P**GIKVRQLCKL	1.45E-07	0.00	R
	B*5702	TDYWQATWIP**E**WEFVNTPPLV	5.49E-05	0.05	
	B*0801	VTDRGRQKIV**S**LTETTNQKTE	8.48E-07	0.00	
	B*18(01)	TNQKTELQAI**Q**LALDSGSEVN	1.50E-05	0.00	
	B*4403	AIQLALQDSG**S**EVNIVTDSQY	2.56E-09	0.00	R
	B*39(10) [Cw*1203][b]	ALQDSGSEVN**I**VTDSQYALGI	3.37E-05	0.05	R
	B*3910[b]	ALGIIQAQPD**K**SESELVNQII	1.81E-05	0.02	R
Integrase	B*4403	LFLDGIDKAQ**E**EHEKYHSNWR	2.28E-05	0.00	R
	B*4403	FLDGIDKAQE**E**HEKYHSNWRA	6.04E-11	0.00	R
	B*4201	AMASEFNLPP**I**VAKEIVASCD	4.65E-06	0.00	R
	B*4201 [Cw*1701]	SEFNLPPIVA**K**EIVASCDKCQ	1.38E-06	0.00	R
	B*42(01) [Cw*1701][b]	EIVASCDKCQ**L**KGEAIHGQVD	3.95E-06	0.01	R
	B*1510	QLDCTHLEGK**V**ILVAVHVASG	1.64E-04	0.05	
	B*4403	SGYIEAEVIP**A**ETGQETAYYI	3.01E-04	0.06	R
	B*5801 [A*02]	HTDNGSNFTS**A**AVKAACWWAG	9.77E-04	0.05	R
	B*1503 [Cw*02]	VKAACWWAGI**Q**QEFGIPYNPQ	2.30E-07	0.00	R
	B*1503 [Cw*02]	MAVFIHNFKR**K**GGIGGYSAGE	2.07E-09	0.00	R

[a] Boldface indicates site of polymorphism (consensus amino acids listed). Underlining indicates sites of known HLA epitopes restricted by the selecting allele.
[b] Optimal epitope not defined.

question of whether the association of particular HLA class I alleles with characteristic levels of viremia relates also to the number of Gag-specific epitopes that are targeted in chronic infection, the HIV-specific $CD8^+$ T-cell responses were first characterized by testing recognition in IFN-γ ELISPOT assays with a panel of peptides spanning the entire C-clade proteome.

Previously, 160 dominant $CD8^+$ T-cell responses and HLA restrictions were defined from a study cohort of 578 subjects (23). We repeated this analysis with the same methodology for the extended cohort of 681 subjects described here. HLA restriction of the ELISPOT response was therefore determined by statistical association, with a Bonferroni correction to take account of multiple tests resulting from the analysis of all 65 HLA class I alleles (23 HLA-A, 25 HLA-B, and 17 HLA-C) expressed in the study cohort at a phenotypic frequency of ≥0.5%. By this approach, an additional 45 significant new associations (in each case, $P < 0.00075$) between the expression of particular class I molecules and ELISPOT peptide recognition were revealed (data not shown), making a total of 205 associations. That the Bonferroni correction is more conservative than the false-discovery rate approach is illustrated by the calculation of the *q* value corresponding to $P < 0.00075$, which is $q < 0.016$. As described previously in the $n = 578$ subset of the cohort, the most immunogenic proteins identified here were Gag (42/205, 20% of the epitopes targeted), Pol (83/205, 41%), and Nef (22/205, 11%); 58% of the responses were HLA-B restricted, 27% were HLA-A restricted, and 13% were HLA-C restricted.

Since HLA-B allele expression has the strongest HLA-mediated influence on the viral set point (22) and the breadth of

TABLE 6. Sites of sequence polymorphism in HIV-1 Nef associated with host HLA-B

Protein	HLA	Polymorphism location[a]	*P* value	*q* value	Reversion
Nef	B*4501	AWLQAQEEEE**E**EGVGFPVRPQ	8.26E-09	0.00	R
	B*07 [Cw*0702]	EEEEGVGFPV**R**PQVPLRPMTY	1.02E-09	0.00	R
	B*8101	VGFPVRPQVP**L**RPMTYKAAFD	1.78E-10	0.00	R
	B*35(01)	RPQVPLRPMT**Y**KAAFDLSFFL	1.13E-07	0.00	R
	B*5801	QVPLRPMTYK**A**AFDLSFFLKE	7.32E-06	0.00	R
	B*4403	AAFDLSFFLK**E**KGGLEGLIYS	2.78E-07	0.00	R
	B*4403	KEKGGLEGLI**Y**SKKRQEILDL	1.89E-08	0.00	R
	B*18(01)	EGLIYSKKRQ**E**ILDLWVYHTQ	1.44E-09	0.00	R
	B*4403	EGLIYSKKRQ**E**ILDLWVYHTQ	1.82E-08	0.00	R
	B*35(01)	DWQNYTPGPG**V**RYPLTFGWCF	6.46E-06	0.00	R
	B*53(01)	DWQNYTPGPG**V**RYPLTFGWCF	1.71E-04	0.02	

[a] Boldface indicates site of polymorphism (consensus amino acids listed). Underlining indicates sites of known HLA epitopes restricted by the selecting allele.

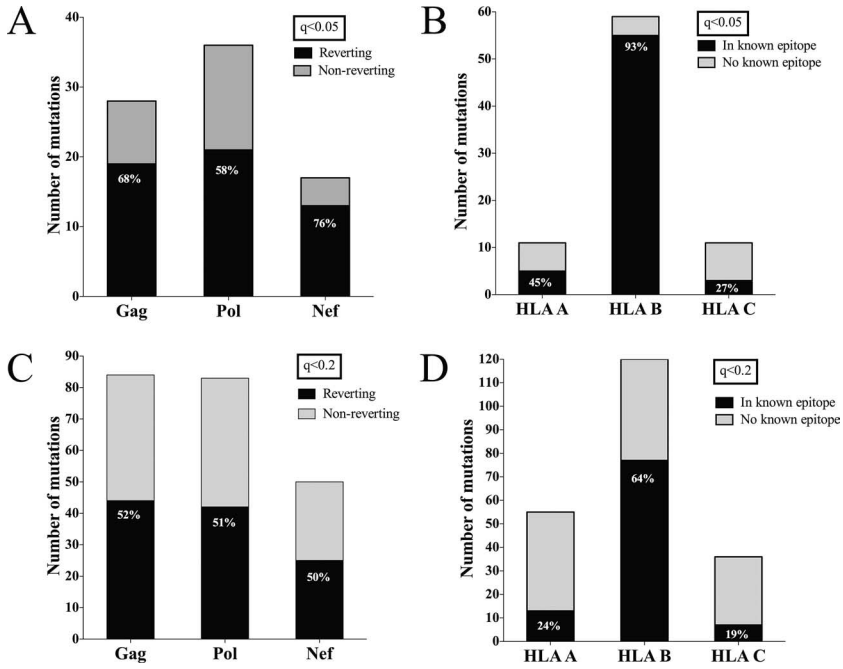

FIG. 1. Number of HIV polymorphisms associated with host HLA, derived from analysis of C-clade sequences from 710 individuals. The number of HLA-associated HIV mutations in Gag, Pol, and Nef and the percentage of reverting sites are shown for $q < 0.05$ (A) and $q < 0.2$ (C). The total numbers of HLA-associated mutations according to HLA class I restriction are shown for associations with $q < 0.05$ (B) and $q < 0.2$ (D).

only Gag-specific $CD8^+$ T-cell responses was associated with a decreasing viral load (23), we sought a correlation between the median viral loads associated with each HLA-B allele in the study population and the number of Gag epitopes presented by that HLA-B allele. A significant association was indeed observed ($r = -0.49$, $P = 0.013$; Fig. 4A). Even excluding the alleles most strongly associated with the control of viremia in this cohort, HLA-B*57 and -B*5801, which are known to present multiple Gag epitopes, the median viral load for each HLA-B allele remains correlated significantly with the number of Gag epitopes presented by that HLA-B allele ($r = -0.50$, $P = 0.018$; data not shown). In contrast, no significant associations were observed for the number of Gag epitopes presented by HLA-A or -C alleles or for epitopes targeted within any non-Gag protein presented by HLA-A, -B, or -C alleles (Fig. 4 B to D; data not shown). These data support earlier findings that, of the HLA class I alleles expressed, HLA-B alleles have the strongest impact on the viral load (22) and that Gag-specific responses in chronic infection are those most strongly associated with the immune control of HIV (23). However, the stronger correlations described above in relation to the number of reverting Gag and Pol polymorphisms associated with different HLA molecules suggest that non-HLA-B alleles and Pol-specific responses can also make important contributions to the immune control of HIV.

DISCUSSION

These analyses of the Gag, Pol, and Nef sequences from 710 HIV-infected subjects point to the critical aspects of the $CD8^+$ T-cell response that are correlated with the immune control of HIV infection. We conclude that the $CD8^+$ T-cell responses driving immune control are those that select escape mutations that inflict a fitness cost upon the virus and that therefore revert posttransmission. Specifically, HLA class I associations with the viral set point that have previously been described are here shown to hinge on the number of reverting mutations selected in Gag and Pol by HLA-B and HLA-C alleles. This correlation is most clear-cut for different HLA-B alleles and the number of reverting mutations in Gag. This association also holds true when using the CD4 count, rather than the viral load, as a marker of disease progression, confirming the central role of reverting escape mutations in HIV disease progression.

Other recent studies have proposed viral fitness cost as a possible mechanism in the control of HIV. Studies of the Gag epitopes restricted by HLA-B*57 and -B*27, the class I molecules most consistently associated with the control of viremia, have demonstrated that the escape mutations selected substantially reduce viral fitness in vitro (28, 31). A study based on published epitopes hypothesized that "effective" $CD8^+$ T-cell responses may be defined by their ability to select escape mutations that impose a fitness cost on the virus (11). Recent studies of early HIV infection support the findings we describe here, indicating that transmission of viruses carrying escape mutations within HLA-B-restricted Gag epitopes is of benefit to the recipient and is associated with lower viral set points (7, 14). Corresponding studies of SIV infection in vaccinated Burmese macaques indicated similarly that durable control is achieved through early escape within a Gag epitope, resulting in a fitness cost to the virus, in combination with multiple additional Gag-specific $CD8^+$ T-cell responses (20, 30). Our study unifies these findings and confirms the central role of mutations that impose a fitness cost (here defined by sites deemed to revert) in CTL-mediated control of viremia.

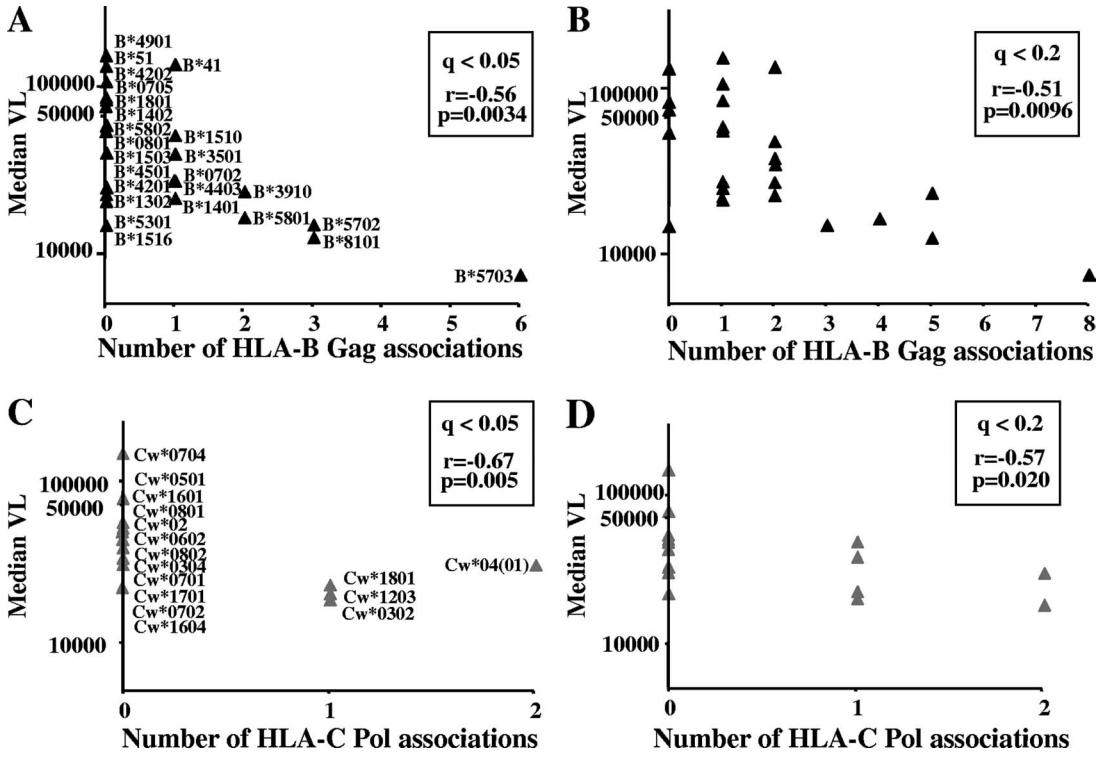

FIG. 2. Relationship between the number of HLA-B-associated polymorphisms in C-clade Gag (n = 673) and Pol (n = 461) sequences and the median viral load. The number of HIV sequence polymorphisms associated with alleles expressed at a ≥0.5% phenotypic frequency is plotted against the median viral load for patients expressing that allele. Shown are HLA-B-selected mutations in Gag for $q < 0.05$ (A) and $q < 0.2$ (B) and HLA-C-selected mutations in Pol for $q < 0.05$ (C) and $q < 0.2$ (D). We excluded associations where a four-digit type was unknown and associations pertaining to more than one allele at a given position (due to linkage), unless the allele responsible for the association could be determined on the basis of a previously defined epitope with the correct restriction.

The most clear-cut correlation we show here is between mutations selected in Gag by HLA-B. Broad Gag-specific $CD8^+$ T-cell responses are associated with good immune control (23), and HLA-B effects have previously been shown to dominate the HLA-mediated influence on disease outcome in HIV infection (22). The new data presented here support both of these findings but also further develop our understanding of the mechanism of these effects by highlighting the central role of mutations that impose a fitness cost and that therefore revert posttransmission. The biological phenomenon of reversion is detected here by statistical inference based on the analysis of a large cross-sectional cohort. The HLA-B associations with the viral set point are significantly correlated with the number of Gag epitopes targeted, even if B*57 and B*5801 are excluded, showing that this finding is not only mediated by the two alleles known to have the strongest effect on the viral load.

In contrast to previous studies, these data also highlight a potential role for mutations selected by HLA-C in Pol. Again, reverting mutations drive the statistical association, highlighting the importance of mutations that impose a cost on viral fitness. The role of HLA-C has historically been liable to underestimation, partly because the limitations of serological HLA typing methods affected HLA-C disproportionately and partly because of the many strong linkage disequilibrium effects between HLA-B and HLA-C. Our study has the advantage of high-resolution HLA typing methods and robust correction for linkage disequilibria and is consistent with the analysis of another large cohort that suggested HLA-C-mediated effects on the suppression of viremia (9).

The observation here that there may be a potential role for HLA-C-restricted targeting of Pol in early infection, evidenced by HLA-C-associated escape in chronic infection, is unexpected, given the previous lack of evidence that HLA-C-restricted responses were associated with improved immune control of HIV, even when Gag was targeted (23). It is possible that, given the small number of reverting HLA-C-associated mutations, the beneficial effect of these responses may be obscured by the larger number of HLA-C-restricted responses that tend not to drive selection pressure on the virus. In contrast, the proportion of mutations associated with HLA-B in Gag, Pol, and Nef (58 [76%] of 76) is somewhat higher than the proportion of $CD8^+$ T-cell responses in these proteins that are associated with HLA-B (84 [63%] of 150; $P = 0.05$ [Fisher's exact test]). Even more striking is the observation of relatively few HLA-A-associated mutations (8 [11%] of 76) for the number of HLA-A-restricted $CD8^+$ T-cell responses to Gag, Pol, and Nef (41 [27%] of 150; $P = 0.004$ [Fisher's exact test]). These comparisons suggest that HLA-B-restricted $CD8^+$ T-cell responses may be more likely to drive selection pressure on HIV than are HLA-A- or HLA-C-restricted responses. However, in the minority of cases where HLA-A- and HLA-C-restricted responses do drive reverting mutations, these are also of benefit in terms of immune control.

The strong correlation between the number of reverting

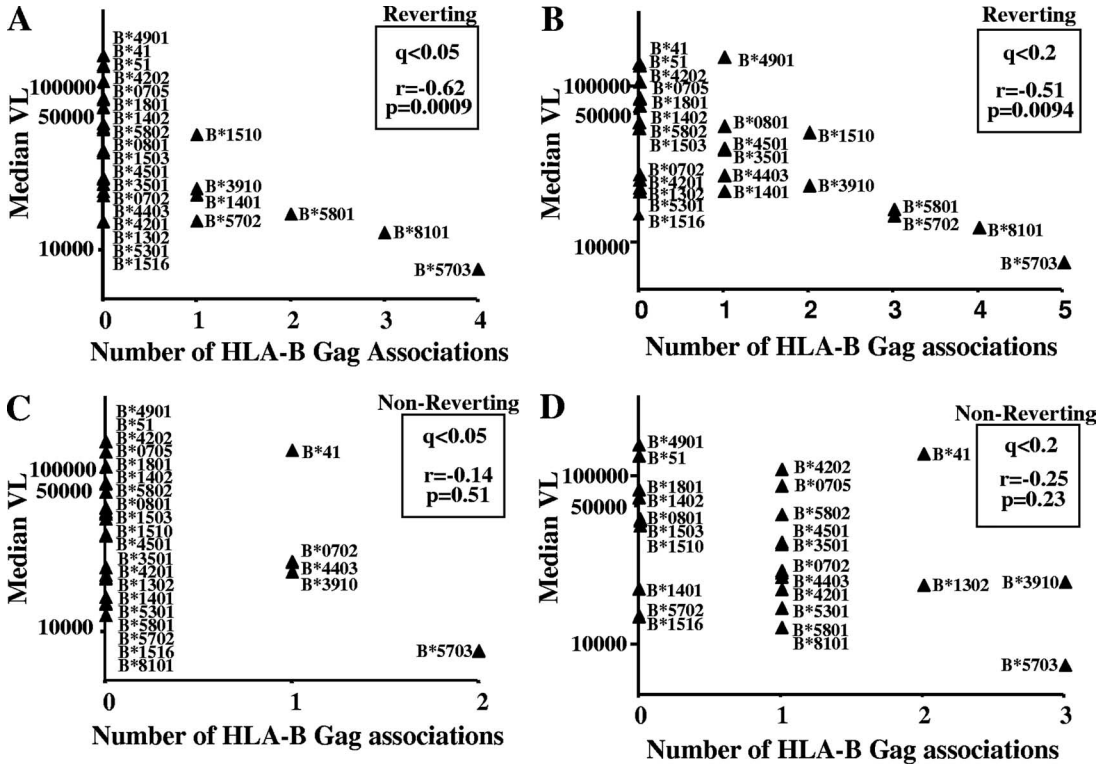

FIG. 3. Relationship between the number of HLA-B-associated polymorphisms in C-clade Gag sequences ($n = 673$) and the median viral load, according to the presence or absence of reversion. The number of sites of HIV polymorphism associated with HLA-B alleles expressed at a ≥0.5% phenotypic frequency is plotted against the median viral load for patients with that allele, at sites where reversion was identified (A, B) and where no reversion was identified (C, D), for associations with $q < 0.05$ (A, C) and $q < 0.2$ (B, D).

mutations and a decreasing viral load may appear somewhat counterintuitive, since escape mutations are selected only if they are of benefit to the virus. However, the benefit may only be significant when balanced against the risk of being eliminated by the $CD8^+$ T-cell response, and several studies now demonstrate the consequences, for both donor and recipient, of escape mutations that reduce viral fitness (7, 8, 14, 42). In contrast to the effect mediated by reverting sites, our data show that "nonreverting" mutations do not contribute substantially to a reduction in the median viral load. These are likely to be escape mutations that do not impose a fitness cost on the virus or where a compensatory mutation alleviates fitness constraints (8, 42). Mutations that do not revert can accumulate in populations and may eventually replace the population wild type (26). The implication of these accumulating mutations is that, first, epitopes are potentially lost over time and no longer contribute to CTL-mediated control and, second, that the escape occurring at these sites does not confer the advantage of reduced viral fitness. Epitope stability may therefore be an important consideration in the selection of potential immunogens for a T-cell vaccine.

The methodology used here was designed to maximize the sensitivity and specificity with which HLA-selected polymorphisms were detected in a large cross-sectional cohort. The statistical approach to the identification of reversion allowed us to test the hypothesis that the selection of mutations that impose a fitness cost is the mechanism underlying the established HLA associations with viral set points. This method has potential limitations; reversion may occur late in the course of disease or following the selection of a rare mutation or may not be detected due to the coexistence of compensatory mutations. In these instances, our methodology may underestimate reverting sites. Furthermore, although viral set points tend to be characteristic of individual HLA alleles, there is still unexplained variance in control of viremia, and clearly, factors other than HLA genotype also have a significant impact.

The mechanisms by which HLA-B and HLA-C specifically mediate their effects through mutations in Gag and Pol, respectively, are unknown. The marked genetic diversity of HLA-B may account for a broad presentation of HIV epitopes, and HLA-B-restricted epitopes are particularly enriched within Gag. Furthermore, in the SIV model, Gag has been shown in vitro to be the first protein presented on the cell surface following acute infection (40). Early targeting of Gag may enable killing of target cells before viral progeny are produced, therefore reducing the viral load. It has been suggested that high expression of HLA-C might be associated with successful control of HIV (9). This prompts the hypothesis that HLA-C alleles selecting reverting escape mutants in Pol might be those expressed at high levels. The emergence of Pol as a potentially important $CD8^+$ T-cell target is also consistent with recent kinetic studies that showed that both Gag- and Pol-specific $CD8^+$ T cells may be activated by virus-infected cells prior to Nef-mediated MHC class I downregulation (41). Other host genetic factors may also contribute to the control of viremia. These may be artificially attributed to HLA-mediated

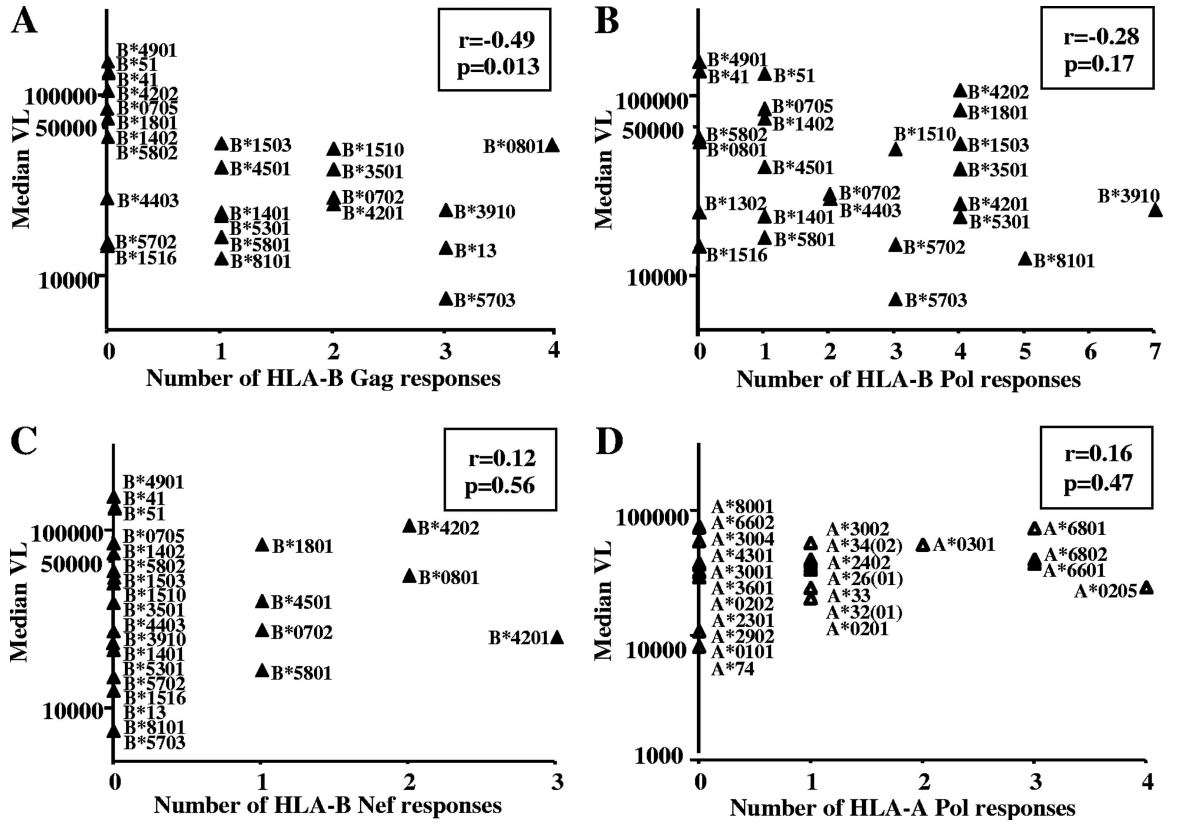

FIG. 4. Relationship between $CD8^+$ T-cell responses determined by IFN-γ ELISPOT assay and the median viral load for 681 HIV-infected individuals. The numbers of statistically significant ($P < 0.00075$) HLA-B-associated responses to peptides in Gag (A), Pol (B), and Nef (C) and HLA-A-associated responses to Pol (D) are plotted against the median viral load for patients expressing that allele.

effects if they are in genetic linkage with the MHC locus. However, a recent genome-wide study identified both HLA-B*57 and HLA-C as the most important genetic determinants of viral control (9).

In view of recent data highlighting the potential role of HLA-C (9), it was essential to use an approach that did not focus on previously characterized epitopes. Thus, we sought associations between any polymorphism arising at any amino acid position within Gag, Pol, or Nef with any HLA class I allele. In this way, we could analyze the role of HLA-C alleles, as well as gain further insight into the patterns of immune control already known to be associated with HLA-B. Critically, the method depended on appropriate correction for multiple comparisons (via the calculation of q values), for sequence relatedness (2), and for linkage disequilibrium effects. Although the phenomena of escape and reversion in this study are based on statistical inference, the validity of our approach is supported by the observation that 93% of HLA-B-associated Gag mutations fall within known epitopes restricted by the appropriate allele (Fig. 1) and by the identification of well-characterized sites of HLA selection pressure.

These data may be relevant in the context of the recent failed Merck vaccine trial (34, 43), although the reasons for the vaccine failure are incompletely understood. One possibility is that the breadth of induced Gag, Pol, and Nef $CD8^+$ T-cell responses in vaccinees (a median of one response per protein) was not sufficient to protect against transmission or to influence the viral set point in vaccinees who subsequently became infected. The implications from this study remain that broad, Gag-specific $CD8^+$ T-cell responses restricted by HLA-B alleles should be induced by a successful vaccine. However, these data also show that targeting of Pol epitopes by HLA-C-restricted $CD8^+$ T cells may contribute to improved control via the same mechanism of selecting escape mutations that inflict a fitness cost on the virus. Although escape by the virus from a single epitope is almost inevitable, the induction of several Gag and Pol responses by a vaccine would ensure that multiple mutations would be needed to bring about escape. The data presented here suggest that these mutations, in themselves, may contribute to immune control via an effect on viral replicative capacity. There remains an urgent need to develop vaccines capable of inducing a more multifaceted Gag- and Pol-specific $CD8^+$ T-cell response in vaccinees than has hitherto been achieved.

ACKNOWLEDGMENTS

This work was funded by the National Institutes of Health (contract NO1-A1-15422, 2RO1AI46995-06, and R01A1067073), the Wellcome Trust (P.G.), the UK Medical Research Council (A.P. and P.M.), and the Mark and Lisa Schwartz Foundation. P.G. is an Elizabeth Glaser Pediatric AIDS Foundation scientist.

REFERENCES

1. **Altfeld, M., M. M. Addo, E. S. Rosenberg, F. M. Hecht, P. K. Lee, M. Vogel, X. G. Yu, R. Draenert, M. N. Johnston, D. Strick, T. M. Allen, M. E. Feeney, J. O. Kahn, R. P. Sekaly, J. A. Levy, J. K. Rockstroh, P. J. Goulder, and B. D. Walker.** 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS **17:**2581–2591.
2. **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.
3. **Borghans, J. A., A. Molgaard, R. J. de Boer, and C. Kesmir.** 2007. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. PLoS ONE **2:**e920.
4. **Brockman, M. A., A. Schneidewind, M. Lahaie, A. Schmidt, T. Miura, I. Desouza, F. Ryvkin, C. A. Derdeyn, S. Allen, E. Hunter, J. Mulenga, P. A. Goepfert, B. D. Walker, and T. M. Allen.** 2007. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J. Virol. **81:**12608–12618.
5. **Bunce, M., C. M. O'Neill, M. C. Barnardo, P. Krausa, M. J. Browning, P. J. Morris, and K. I. Welsh.** 1995. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens **46:**355–367.
6. **Carlson, J., C. Kadie, S. Mallal, and D. Heckerman.** 2007. Leveraging hierarchical population structure in discrete association studies. PLoS ONE **2:**e591.
7. **Chopera, D. R., Z. Woodman, K. Mlisana, M. Mlotshwa, D. P. Martin, C. Seoighe, F. Treurnicht, D. A. de Rosa, W. Hide, S. A. Karim, C. M. Gray, and C. Williamson.** 2008. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog. **4:**e1000033.
8. **Crawford, H., J. G. Prado, A. Leslie, S. Hué, I. Honeyborne, S. Reddy, M. van der Stok, Z. Mncube, C. Brander, C. Rousseau, J. I. Mullins, R. Kaslow, P. Goepfert, S. Allen, E. Hunter, J. Mulenga, P. Kiepiela, B. D. Walker, and P. J. Goulder.** 2007. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. J. Virol. **81:**8346–8351.
9. **Fellay, J., K. V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J. M. Miro, N. Obel, J. P. Smith, J. Wyniger, P. Descombes, S. E. Antonarakis, N. L. Letvin, A. J. McMichael, B. F. Haynes, A. Telenti, and D. B. Goldstein.** 2007. A whole-genome association study of major determinants for host control of HIV-1. Science **317:**944–947.
10. **Fernandez, C. S., I. Stratov, R. De Rose, K. Walsh, C. J. Dale, M. Z. Smith, M. B. Agy, S.-L. Hu, K. Krebs, D. I. Watkins, D. H. O'Connor, M. P. Davenport, and S. J. Kent.** 2005. Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost. J. Virol. **79:**5721–5731.
11. **Frater, A. J., H. Brown, A. Oxenius, H. F. Gunthard, B. Hirschel, N. Robinson, A. J. Leslie, R. Payne, H. Crawford, A. Prendergast, C. Brander, P. Kiepiela, B. D. Walker, P. J. Goulder, A. McLean, and R. E. Phillips.** 2007. Effective T-cell responses select human immunodeficiency virus type 1 mutants and slow disease progression. J. Virol. **81:**6742–6751.
12. **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.
13. **Geldmacher, C., J. R. Currier, E. Herrmann, A. Haule, E. Kuta, F. McCutchan, L. Njovu, S. Geis, O. Hoffmann, L. Maboko, C. Williamson, D. Birx, A. Meyerhans, J. Cox, and M. Hoelscher.** 2007. CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients. J. Virol. **81:**2440–2448.
14. **Goepfert, P. A., W. Lumm, P. Farmer, P. Matthews, A. Prendergast, J. M. Carlson, C. A. Derdeyn, J. Tang, R. A. Kaslow, A. Bansal, K. Yusim, D. Heckerman, J. Mulenga, S. Allen, P. J. Goulder, and E. Hunter.** 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J. Exp. Med. **205:**1009–1017.
15. **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.
16. **Goulder, P. J., M. Bunce, G. Luzzi, R. E. Phillips, and A. J. McMichael.** 1997. Potential underestimation of HLA-C-restricted cytotoxic T-lymphocyte responses. AIDS **11:**1884–1886.
17. **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.
18. **Goulder, P. J., and D. I. Watkins.** 2004. HIV and SIV CTL escape: implications for vaccine design. Nat. Rev. Immunol. **4:**630–640.
19. **Kaslow, R. A., M. Carrington, R. Apple, L. Park, A. Munoz, A. J. Saah, J. J. Goedert, C. Winkler, S. J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich, and D. L. Mann.** 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat. Med. **2:**405–411.
20. **Kawada, M., H. Igarashi, A. Takeda, T. Tsukamoto, H. Yamamoto, S. Dohki, M. Takiguchi, and T. Matano.** 2006. Involvement of multiple epitope-specific cytotoxic T-lymphocyte responses in vaccine-based control of simian immunodeficiency virus replication in rhesus macaques. J. Virol. **80:**1949–1958.
21. **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.
22. **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.
23. **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.
24. **Klein, M. R., C. A. van Baalen, A. M. Holwerda, S. R. Kerkhof Garde, R. J. Bende, I. P. Keet, J. K. Eeftinck-Schattenkerk, A. D. Osterhaus, H. Schuitemaker, and F. Miedema.** 1995. Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. J. Exp. Med. **181:**1365–1372.
25. **Kobayashi, M., H. Igarashi, A. Takeda, M. Kato, and T. Matano.** 2005. Reversion in vivo after inoculation of a molecular proviral DNA clone of simian immunodeficiency virus with a cytotoxic-T-lymphocyte escape mutation. J. Virol. **79:**11529–11532.
26. **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.
27. **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.
28. **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.
29. **Masemola, A., T. Mashishi, G. Khoury, P. Mohube, P. Mokgotho, 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, and C. M. Gray.** 2004. Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by $CD8^+$ T cells: correlation with viral load. J. Virol. **78:**3233–3243.
30. **Matano, T., M. Kobayashi, H. Igarashi, A. Takeda, H. Nakamura, M. Kano, C. Sugimoto, K. Mori, A. Iida, T. Hirata, M. Hasegawa, T. Yuasa, M. Miyazawa, Y. Takahashi, M. Yasunami, A. Kimura, D. H. O'Connor, D. I. Watkins, and Y. Nagai.** 2004. Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. J. Exp. Med. **199:**1709–1718.
31. **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.

32. **Migueles, S. A., M. S. Sabbaghian, W. L. Shupert, M. P. Bettinotti, F. M. Marincola, L. Martino, C. W. Hallahan, S. M. Selig, D. Schwartz, J. Sullivan, and M. Connors.** 2000. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc. Natl. Acad. Sci. USA **97:**2709–2714.
33. **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.
34. **Moore, J. P., P. J. Klasse, M. J. Dolan, and S. K. Ahuja.** 2008. AIDS/HIV. A STEP into darkness or light? Science **320:**753–755.
35. **Nixon, D. F., A. R. Townsend, J. G. Elvin, C. R. Rizza, J. Gallwey, and A. J. McMichael.** 1988. HIV-1 gag-specific cytotoxic T lymphocytes defined with recombinant vaccinia virus and synthetic peptides. Nature **336:**484–487.
36. **O'Brien, S. J., X. Gao, and M. Carrington.** 2001. HLA and AIDS: a cautionary tale. Trends Mol. Med. **7:**379–381.
37. **Ogg, G. S., X. Jin, S. Bonhoeffer, P. R. Dunbar, M. A. Nowak, S. Monard, J. P. Segal, Y. Cao, S. L. Rowland-Jones, V. Cerundolo, A. Hurley, M. Markowitz, D. D. Ho, D. F. Nixon, and A. J. McMichael.** 1998. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. Science **279:**2103–2106.
38. **Rivière, Y., M. B. McChesney, F. Porrot, F. Tanneau-Salvadori, P. Sansonetti, O. Lopez, G. Pialoux, V. Feuillie, M. Mollereau, S. Chamaret, et al.** 1995. Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. AIDS Res. Hum. Retrovir. **11:**903–907.
39. **Rousseau, C. M., B. A. Birditt, A. R. McKay, J. N. Stoddard, T. C. Lee, S. McLaughlin, S. W. Moore, N. Shindo, G. H. Learn, B. T. Korber, C. Brander, P. J. Goulder, P. Kiepiela, B. D. Walker, and J. I. Mullins.** 2006. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. J. Virol. Methods **136:**118–125.
40. **Sacha, J. B., C. Chung, E. G. Rakasz, S. P. Spencer, A. K. Jonas, A. T. Bean, W. Lee, B. J. Burwitz, J. J. Stephany, J. T. Loffredo, D. B. Allison, S. Adnan, A. Hoji, N. A. Wilson, T. C. Friedrich, J. D. Lifson, O. O. Yang, and D. I. Watkins.** 2007. Gag-specific $CD8^+$ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. J. Immunol. **178:**2746–2754.
41. **Sacha, J. B., C. Chung, J. Reed, A. K. Jonas, A. T. Bean, S. P. Spencer, W. Lee, L. Vojnov, R. Rudersdorf, T. C. Friedrich, N. A. Wilson, J. D. Lifson, and D. I. Watkins.** 2007. Pol-specific $CD8^+$ T cells recognize simian immunodeficiency virus-infected cells prior to Nef-mediated major histocompatibility complex class I downregulation. J. Virol. **81:**11703–11712.
42. **Schneidewind, A., M. A. Brockman, R. Yang, R. I. Adam, B. Li, S. Le Gall, C. R. Rinaldo, S. L. Craggs, R. L. Allgaier, K. A. Power, T. Kuntzen, C. S. Tung, M. X. LaBute, S. M. Mueller, T. Harrer, A. J. McMichael, P. J. Goulder, C. Aiken, C. Brander, A. D. Kelleher, and T. M. Allen.** 2007. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J. Virol. **81:**12382–12393.
43. **Steinbrook, R.** 2007. One step forward, two steps back—will there ever be an AIDS vaccine? N. Engl. J. Med. **357:**2653–2655.
44. **Zuñiga, R., A. Lucchetti, P. Galvan, S. Sanchez, C. Sanchez, A. Hernandez, H. Sanchez, N. Frahm, C. H. Linde, H. S. Hewitt, W. Hildebrand, M. Altfeld, T. M. Allen, B. D. Walker, B. T. Korber, T. Leitner, J. Sanchez, and C. Brander.** 2006. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. J. Virol. **80:**3122–3125.