NIH Public Access
Author Manuscript
AIDS. Author manuscript; available in PMC 2010 March 13.

Published in final edited form as:
AIDS. 2009 March 13; 23(5): 579–587. doi:10.1097/QAD.0b013e328328f76e.

HIV-1 evolution in *gag* and *env* is highly correlated but exhibits different relationships with viral load and the immune response

Anne Piantadosi[a,b], **Bhavna Chohan**[a,b], **Dana Panteleeff**[a], **Jared M. Baeten**[c], **Kishorchandra Mandaliya**[d], **Jeckoniah O. Ndinya-Achola**[e], and **Julie Overbaugh**[a,b]

[a] Division of Human Biology, Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, USA

[b] Graduate Program in Pathobiology, University of Washington, Seattle, Washington, USA

[c] Department of Global Health, University of Washington, Seattle, Washington, USA

[d] Coast Provincial General Hospital, Mombasa

[e] Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya

Abstract

Objective—To evaluate relationships between HIV-1 evolution, including immune evasion, and markers of disease progression during chronic infection.

Design—HIV-1 evolution and disease progression markers were evaluated over approximately 5 years of infection among 37 Kenyan women from a prospective, seroincident cohort. Evolution was measured in two genes, *gag* and *env*, which are primary targets of cellular and humoral immune responses, respectively.

Methods—Proviral HIV-1 *gag* and *env* sequences were obtained from early and chronic infection when plasma viral load and CD4 cell counts were available. Human leukocyte antigen types were obtained to identify changes in *gag* cytotoxic T lymphocyte epitopes. The breadth of the neutralizing antibody response was measured for each woman's plasma against a panel of six viruses. Tests of association were performed between virus evolution (diversity, divergence, and ratio of nonsynonymous to synonymous divergence), markers of disease progression (viral load and CD4 cell count), and immune parameters (*gag* cytotoxic T lymphocyte epitope mutation and neutralizing antibody breadth).

Results—HIV-1 *gag* and *env* diversity and divergence were highly correlated in early and late infection. Divergence in *gag* was strongly correlated with viral load, largely because of the accumulation of synonymous changes. Mutation in *gag* cytotoxic T lymphocyte epitopes was associated with higher viral load. There was evidence for adaptive evolution in *env*, but the extent of *env* evolution was only weakly associated with neutralizing antibody breadth.

Conclusion—Our results indicate that HIV-1 evolution in *gag* and *env* is highly correlated but exhibits gene-specific differences. The different immune pressures on these genes may partly explain differences in evolution and consequences for HIV-1 disease progression.

Correspondence to Julie Overbaugh, Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail stop C3-168, Seattle, WA 98109-1024, USA E-mail: joverbau@fhcrc.org.

Authors' contributions: Anne Piantadosi performed experiments, analyzed the data, and wrote the manuscript. Bhavna Chohan performed experiments and helped analyze the data. Dana Panteleeff performed neutralization experiments and analysis. Jared M. Baeten helped analyze the data and provided critical input on the manuscript. Kishorchandra Mandaliya and Jeckoniah O. Ndinya-Achola oversaw the clinical and laboratory aspects of this study related to the cohort. Julie Overbaugh oversaw all aspects of this research project, including experimental design, analysis, and writing of the manuscript.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Keywords

cytotoxic T lymphocytes; disease progression; genetic diversity; HIV

Introduction

The HIV-1 is notable for its vast ability to evolve as a result of its error-prone reverse transcriptase and high rate of virus production. Within-host HIV-1 evolution is characterized by diversification of the infecting virus population throughout the course of infection [1]. In addition to genetic drift and purifying selection, HIV-1 can evolve under positive selection from antibodies and cytotoxic T lymphocytes (CTLs) [1]. However, it remains unclear to what extent virus evolution and escape from immune pressure contribute to HIV-1 disease progression.

Studies examining whether the extent of HIV-1 evolution is associated with disease progression have had conflicting results. HIV-1 evolution is most commonly evaluated by measuring viral diversity and divergence in nucleotide sequences as well as indicators of selection such as the ratio of nonsynonymous to synonymous change (dN/dS). Shankarappa *et al.* [2] described a consistent pattern of increasing virus diversity and divergence throughout infection, which was temporally associated with progression to AIDS. In concordance with this, Markham *et al.* [3] found greater HIV-1 diversity and divergence, as well as evidence of adaptive evolution, in rapid progressors compared with moderate or slow progressors. By contrast, Ganeshan *et al.* [4] found greater diversity and stronger evidence of adaptive evolution in slow progressors than in rapid progressors. The conflicting results of these and other studies, which included a relatively small number of participants (≤15) [5–8], highlight the need for larger studies to assess the relationship between HIV-1 evolution and disease progression.

The relationship between HIV-1 evolution and disease progression remains particularly unclear with respect to an important component of HIV-1 evolution, evasion of the adaptive immune response. For example, some individual CTL escape mutations are associated with disease progression [9–12], whereas others are not because they impose a fitness cost on the virus [13–17]. Similarly, there is strong evidence that HIV-1 evolves to escape from neutralizing antibodies (NAbs) throughout infection [18,19], but it is not clear whether this leads to disease progression. One study found an inverse relationship between viral load and the level of autologous neutralization in cross-sectional analysis [20], suggesting that NAb escape facilitates disease progression. However, other studies have demonstrated that longitudinal escape from NAb is not a correlate of disease progression [21,22].

Many previous studies of HIV-1 evolution, disease progression, and the immune response have been limited by cross-sectional study designs, small number of participants, small number of sequences studied per participant, or all. In addition, prior studies have generally focused on one HIV-1 gene and one immune parameter. Here, we present analyses of HIV-1 evolution among 37 Kenyan women followed for an average of 5 years after infection. The primary goal of this study was to evaluate the relationship between HIV-1 evolution and disease progression and to compare this relationship in two different genes, *gag* and *env*. Because these genes are primary targets of different arms of the immune response (CTLs and NAb, respectively), immune evasion was also evaluated in relation to virus evolution. This is the first study to directly compare the potential role of both immune factors in driving HIV-1 evolution and disease progression.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Methods

Study population

Individuals in this study were part of a prospective seroincident cohort of high-risk Kenyan women [23]. Methods for determining the timing of HIV-1 infection and measuring plasma viral load and CD4 cell count have been described previously [24]. Viral load set point was defined as the first available measurement 4–24 months after infection [24]. Thirty-seven women were included in this study, which was restricted to those who had a blood sample taken within the first year of infection, were antiretroviral naïve at the time of sample collection, and were not coinfected or superinfected [25,26]. In addition, all women in this study had subtype A viruses based on *env* V1–V3 sequences [27]; however, *gag* sequencing ultimately revealed that six women were initially infected with recombinant viruses containing subtype D in *gag* (described in results). The study was approved by the ethical review committees of the University of Nairobi, the University of Washington, and the Fred Hutchinson Cancer Research Center.

Sequence analysis

HIV-1 sequences were obtained as described previously [26,28]. Briefly, proviral DNA was extracted from frozen peripheral blood mononuclear cells (PBMCs) using the QIAamp DNA Blood Mini kit (Qiagen, Valencia, California, USA), and the HIV-1 proviral copy number was estimated using quantitative PCR [29,30]. *Envelope* V1–V5 (~1.2 kb) and *gag* p17 partial p24 (~700 bp) were amplified in separate nested PCRs, each from an estimated single copy of proviral template, using previously published primers [26]. PCR products were sequenced directly. A median of eight *gag* sequences (range, 3–15) and seven *env* sequences (range, 3–13) were obtained from each sample. *Gag* and *env* V1–V3 sequences from seven of the 37 individuals were published previously [28].

Sequences were assembled using Sequencher (Gene Codes, Ann Arbor, Michigan, USA) and aligned using MacClade 4.0 [31]. Regions that could not be unambiguously aligned between different individuals (mainly regions of *env* V1/V2 and V4) were removed from the analysis. Although this likely led to an underestimation of the *env* diversity within each individual, it allowed us to compare the same region of the gene across all individuals. For each individual, the most recent common ancestor (MRCA) of the sequences from the early sample was reconstructed using maximum likelihood in GARLI 0.95 [32]. Genetic distances were calculated using a general time reversible model ($\gamma = 0.5$) in phylogenetic analysis using parsimony [33]. Synonymous and nonsynonymous substitution rates were determined using Synonymous Non-synonymous Analysis Program (www.hiv.lanl.gov) [34]. Amino acid changes were defined as positions at which two or more of the chronic infection sequences shared an amino acid that was different from the MRCA.

Mutations in gag cytotoxic T lymphocyte epitopes

Each individual's human leukocyte antigen (HLA)-A, B, and C types were determined using an intermediate resolution reverse probe typing system (LabType, One Lambda, Canoga Park, California, USA) in collaboration with the Clinical Immunogenetics Laboratory at the Seattle Cancer Care Alliance. LabType employs locus-specific primers to amplify exons 2 and 3 of the known alleles for the HLA-A, B, and C loci. Amplified sample DNA is hybridized with panels of sequence-specific oligonucleotide probes bound to color-coded microspheres. Probe reactions are captured with a Luminex flow analyzer, and an analysis program compares the pattern of positive and negative probe reactions with a library of known HLA alleles and assigns the typing result.

NIH-PA Author Manuscript

For each allele, published CTL epitopes were identified using the *HIV Molecular Immunology* database (www.lanl.gov). For each individual, we calculated both the total number of amino acid changes in CTL epitopes and the percentage of the total epitopes that contained at least one amino acid change (‘epitopes changed’).

Measurement of neutralizing antibody breadth

Plasma from each individual was tested against a panel comprised of SF162, an easy to neutralize subtype B virus, as well as five subtype-A viruses, Q461d1, Q168b23, Q769h5, Q842d16, and Q259d2.17. These subtype A viruses, which were isolated from other individuals in the Mombasa cohort early in infection [35], were selected because they represent a range of neutralization sensitivities [36]. The median pair-wise difference between these sequences is 13.7% (range, 10.2–18.8%) across the V2–V5 region, and they are further described in [36]. The TZM-bl neutralization assay used for this study has been described [36,37]. Each plasma virus combination was tested in duplicate using plasma dilutions of 1: 100–1: 1600, and the 50% maximal inhibitory concentration (IC_{50}) was calculated from the mean of the duplicate results as the reciprocal plasma dilution required for 50% inhibition. The median IC_{50} for each virus against all plasma samples was determined, and the NAb breadth score for each plasma was calculated as the number of panel viruses that were neutralized to greater than the median for that experiment. The breadth scores reported are the mean of three independent experiments.

Statistical analyses

Associations between diversity, divergence, plasma viral load, CD4 cell count, and immune parameters were assessed using the nonparametric Spearman’s rank correlation test for univariate comparisons and linear regression for multivariate analyses. The Wilcoxon signed-rank test was used to compare various measurements of evolution between genes. All analyses were performed in STATA 9 (Stata, College Station, Texas, USA) [38].

Results

Characterization of HIV-1 evolution in *gag* and *env*

To measure HIV-1 evolution within infected individuals, we examined a total of 651 *gag* and 542 *env* sequences in a longitudinal study of 37 Kenyan women. Proviral sequences were sampled for two times in infection: an ‘early time point,’ which was a median of 88 days after infection (range, 16–299), and a ‘chronic infection time point,’ which was a median of 5.3 years after infection (range, 2.0–7.2). Phylogenetic analysis and genotyping indicated that all *env* sequences belonged to subtype A, which was a selection criterion for this study, whereas in *gag*, we detected subtype A (31 individuals), subtype D (two individuals), and A/D recombinant (four individuals).

For each individual, we calculated the HIV-1 sequence diversity as the mean pair-wise distance between sequences at a given time point. As shown in Fig. 1, the diversity in each gene was lower in early infection than in chronic infection, and *gag* diversity was lower than *env* diversity at both time points. The median HIV-1 diversity among all individuals at the early time point was 0.2% in *gag* and 0.3% in *env*. At the chronic infection time point, the median HIV-1 diversity was 1.1% in *gag* and 2.7% in *env*. To measure the extent of evolution from early to chronic infection, we reconstructed the MRCA of each individual’s early sequences and calculated divergence as the mean distance from a chronic infection sequence to the MRCA. As with diversity, divergence was lower for *gag* (median, 1.7%) than for *env* (median, 3.1%) (Fig. 1). Thus, the average divergence was 0.3% per year in *gag* and 0.6% per year in *env*. It should be noted that the *env* diversity and divergence reported here are likely an underestimate

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of these values across all of V1–V5 because certain highly variable regions were excluded from analysis in order to compare the same region across all individuals.

We found strong correlations between HIV-1 evolution in these two genes as well as across these two times in infection. As shown in Fig. 1, there was a high correlation between *gag* diversity and *env* diversity at both time points as well as between *gag* divergence and *env* divergence. Furthermore, within each gene, there was a high correlation between measures of evolution in early and chronic infection. Thus, individuals who had a high diversity in early infection continued to have a high diversity in chronic infection as well as high divergence ($P \leq 0.01$ for all comparisons).

To measure selection in these genes, we compared synonymous (dS) and nonsynonymous (dN) divergence from the MRCA as well as the dN/dS ratio. As shown in Table 1, we observed comparable levels of synonymous divergence in *gag* and *env*; however, nonsynonymous divergence was significantly lower in *gag* than in *env*. Consequently, the dN/dS ratio was significantly lower in *gag* than in *env* (median dN/dS, 0.4 for *gag* and 1.5 for *env;* $P < 0.001$). As indicated in Table 1, the dN/dS ratio in *gag* was less than or equal to one in all individuals, indicating overall purifying selection. By contrast, the *env* dN/dS ratio was higher than one for 75% of these individuals, and the average *env* dN/dS ratio among all individuals was statistically significantly higher than one (*t*-test, $P < 0.001$), indicating overall positive selection in this gene. In addition to comparing evolution at the nucleotide level, we also examined amino acid changes and found a significantly higher percentage of amino acids changed in *env* than in *gag* (Table 1). Together, these results demonstrate that, although evolution between these genes is highly correlated, there is increased evolution and positive selection in *env* than in *gag*.

Associations between virus evolution and viral load

To evaluate whether evolution in *gag* or in *env* or in both was associated with measures of disease progression during chronic infection, plasma viral load was examined at the chronic infection time point and CD4 cell count at or within 6 months of the chronic infection time point. The median plasma viral load was 5.1 $\log_{10}$ copies/ml (range, 2.6–6.9, available for 33 women), and the median CD4 cell count was 382 cells/μl (range 37–1088, available for 28 women). Greater divergence in *gag*, but not in *env*, was associated with a higher viral load at the chronic infection time point [$\rho = 0.45$, $P = 0.009$ for *gag* (Fig. 2a); $\rho = 0.25$, $P = 0.2$ for *env* (Fig. 2b)]. Divergence in *gag* but not in *env* was highly inversely correlated with CD4 cell count in chronic infection (Fig. 2c and d). Because viral load during chronic infection could be affected by time since infection and viral load set point, we performed multivariate linear regression analysis adjusting for these factors. We found that *gag* divergence remained significantly associated with viral load at the chronic infection time point. Each 1% increase in *gag* divergence was associated with a 0.69 $\log_{10}$ copies/ml higher viral load ($P = 0.01$).

An association between HIV-1 evolution and viral load could reflect either selection for viruses of high fitness or the accumulation of random changes through many rounds of virus replication. Therefore, we independently examined synonymous and nonsynonymous divergence in *gag*. Viral load was highly correlated with *gag* dS ($\rho = 0.56$, $P < 0.001$) but was only marginally, and not statistically significantly, associated with *gag* dN ($\rho = 0.32$, $P = 0.07$). Thus, synonymous changes are likely to contribute significantly to the association that we observed between *gag* divergence and viral load, although we could not exclude a small contribution by nonsynonymous changes.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Mutations in cytotoxic T lymphocyte epitopes

To further evaluate whether the nonsynonymous divergence observed in *gag* might reflect functional changes that could contribute to higher viral load, we evaluated amino acid changes that occurred in published CTL epitopes (www.hiv.lanl.gov). For this purpose, each individual's HLA-A, B, and C alleles were determined to intermediate resolution. We observed a total of 12 different HLA-A alleles, 17 different HLA-B alleles, and 10 different HLA-C alleles. The alleles that we detected at high frequency (15% or higher) were A02, A30, C06, and C07, which have also been reported at high frequency in other African cohorts [39–41].

For each HLA allele, published CTL epitopes were identified using the *LANL HIV Immunology* database. There were no published HIV-1 CTL epitopes for 13 alleles that we detected, although only one of these (C07) was present at high frequency in our cohort. For each individual, we calculated the total number of amino acid changes that occurred in published *gag* CTL epitopes. Individuals had a median of two amino acid changes in published *gag* CTL epitopes (range, 0–5), representing 30% (0–100) of the total number of amino acid changes in *gag*. However, there was no association between the number of amino acid changes in *gag* CTL epitopes and viral load (Fig. 3a).

For each individual, we also calculated the percentage of *gag* epitopes that contained at least one change ('epitopes changed'). Individuals had a median of 27% of epitopes changed (range, 0–100). Individuals with a higher percentage of epitopes changed had higher viral loads during chronic infection ($\rho = 0.44$, $P = 0.01$) (Fig. 3b). This association remained significant in multivariate linear regression analysis controlling for time since infection and viral load set point. For every additional 10% of epitopes changed, viral load was 0.15 $\log_{10}$ copies/ml higher ($P = 0.04$). Therefore, mutations in *gag* CTL epitopes represent another factor contributing to the observed relationship between *gag* evolution and viral load during chronic infection.

Neutralizing antibody breadth

Whereas *gag* evolution was most likely to have been driven by CTL pressure, we hypothesized that evolution in *env* may have been modulated by the NAb response. To address this, the breadth of the NAb response in each individual was measured at or within 1 year of the chronic infection time point (median, 4.9 years after infection; range, 2.0–6.8). The NAb breadth of each plasma sample was calculated as the number of panel viruses (out of six) that were neutralized to greater than the median for that virus, to control for differences in neutralization susceptibility between viruses (Fig. 4). The median NAb breadth score (out of a possible 6) was 2 (range, 0–5.3). NAb breadth was not associated with viral load during chronic infection ($\rho = 0.29$, $P = 0.12$). NAb breadth was also not associated with *env* diversity or divergence during chronic infection ($P > 0.2$ for both). There was a trend for an association between NAb breadth and percentage amino acid change in *envelope* ($\rho = 0.31$, $P = 0.07$). Together, these data suggest that NAb breadth, as measured by neutralization of heterologous viruses, is not a dominant selective force on *env* evolution or virus replication as measured by viral load.

Discussion

We evaluated the relationship between HIV-1 evolution and disease progression in a longitudinal study of 37 Kenyan women. The strengths of this study include the relatively large study population and the examination of many sequences from two genes, which are targets of different adaptive immune responses, at two times in infection. Evolution was highly correlated between *gag* and *env* as well as between early and chronic infection, suggesting that the diversity of the initial virus population influences subsequent evolution. Although there were strong associations between *gag* and *env* evolution, *env* had greater diversity at both time points and greater divergence from early to chronic infection.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Further examination of evolution in these genes revealed that there was no significant difference in synonymous divergence between *gag* and *env*. This suggests that the correlation between evolution in these genes may be largely due to the accumulation of random changes. This interpretation is supported by the association between synonymous divergence in *gag* and viral load; viruses that replicate more will generate both higher viral loads and more synonymous changes. It is interesting that we did not find a relationship between synonymous evolution in *env* and viral load as in the study by Lemey *et al.* [42]. A possible explanation is that several individuals in our study had high early *env* diversity (up to 6.4%), which is not unusual for women who acquire HIV-1 through heterosexual transmission [43,44]. High early diversity could disrupt the relationship between synonymous divergence and viral load because it introduces evolution that occurred in the transmitting partner. Importantly, all of the early sequences from each individual were monophyletic, indicating that there was no evidence for coinfection with two different HIV-1 strains in any individual [26]. Two individuals had notably high early diversity in *env* (>2%); however, removing them from analyses did not significantly alter the results (not shown).

In addition to finding greater diversity and divergence in *env* than in *gag*, we also found evidence for overall positive selection in *env* and purifying selection in *gag*. Because NAbs represent a potential selective force acting on *env* but not on *gag*, we evaluated the breadth of the NAb response in these individuals against a panel of unrelated viruses. There was substantial variation in heterologous NAb breadth, ranging from no detectable neutralization to neutralization of five out of six panel viruses at high levels. Although the TZM-bl neutralization assay may have limitations as compared with the 'gold standard' PBMC assay [45,46], this highly reproducible approach allowed us to compare NAb breadth across individuals in a standardized assay. Individuals who mounted a broader NAb response had more amino acid changes in *env*, although this did not achieve statistical significance (P = 0.07). This result is consistent with a prior study in which *env* amino acid changes were associated with escape from autologous NAb [22]. In this study, we did not assess autologous NAb, which would provide more direct information about the selective pressure faced by these viruses *in vivo*. Nevertheless, our results are consistent with a model proposed by Frost *et al.* [22], in which NAbs exert 'soft' selective pressure that favors the growth of some virus variants over others but does not limit overall growth of the virus population. In concordance with this, we did not observe an association between virus replication, as measured by viral load, and either NAb breadth or *env* diversity or divergence.

In contrast to NAbs, CTLs have been proposed to exert 'hard' selective pressure that limits the overall virus population size [22]; therefore, escape from CTLs should allow increased virus replication. Supporting this model, we found that individuals with a higher percentage change of *gag* CTL epitopes had higher viral loads in chronic infection. A similar relationship between viral load and percentage change of CTL epitopes was observed in a large cross-sectional study [47]. Confirmation of this association by our longitudinal study strengthens the idea that CTL escape may generally contribute to disease progression. A limitation of our study is that we did not perform in-vitro experiments to confirm CTL recognition or escape. However, previous studies have demonstrated associations between mutations in CTL epitopes and in-vitro escape [47–49].

Another limitation of this study is that our estimates of HIV-1 evolution are based on proviral HIV-1 sequences, some of which may be defective. However, other studies have shown that overall patterns of HIV-1 evolution between DNA and RNA are similar [2,50], suggesting that our estimates may be relevant to both sequence types. Importantly, another study demonstrated that the predominant variants found in proviral sequences were maintained after culture, indicating that they are replication-competent [51].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Our study is the first to directly compare longitudinal relationships between HIV-1 evolution in two genes, immune evasion, and disease progression. Our results support a model in which HIV-1 *gag* and *env* evolve in a concerted manner, largely because of similar rates of synonymous evolution. However, because these genes are the primary targets of different adaptive immune responses, gene-specific evolution also occurs. Adaptive evolution in *gag* is relatively rare but may promote disease progression via CTL escape, whereas adaptive evolution in *env* is common and may reflect escape from NAb but does not significantly affect viral load. Further longitudinal studies are needed to more clearly elucidate the selective forces acting on *env*, including autologous NAb. Ultimately, a better understanding of how HIV-1 evolves in response to immune pressure, and how this contributes to disease progression, will provide insight into achieving immune-mediated control of HIV-1 infection.

Acknowledgments

We thank the women who participated in this study. We also thank the Mombasa clinic and laboratory staff for technical assistance, Zahra Jalalian-Lechak for assistance with HLA typing, Anajane G. Smith, Randy Williams, and Rita Witherspoon at the Seattle Cancer Care Alliance Clinical Immunogenetics Laboratory for guidance in HLA typing, and Catherine Blish, Gael Kurath, and Stephen Maley for comments on the manuscript. This work was supported by National Institutes of Health grant A138518. Anne Piantadosi was supported by NIH training grants T32 A107140 and T32 GM07266 and Bhavna Chohan by D43-00007 through the University of Washington International AIDS Research and Training Program.

References

1. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet 2004;5:52–61. [PubMed: 14708016]
2. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 1999;73:10489–10502. [PubMed: 10559367]
3. Markham RB, Wang WC, Weisstein AE, Wang Z, Munoz A, Templeton A, et al. Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc Natl Acad Sci U S A 1998;95:12568–12573. [PubMed: 9770526]
4. Ganeshan S, Dickover RE, Korber BT, Bryson YJ, Wolinsky SM. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. J Virol 1997;71:663–677. [PubMed: 8985398]
5. Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, Whetsell AJ, et al. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science 1996;272:537–542. [PubMed: 8614801]
6. Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol 2002;76:11715–11720. [PubMed: 12388731]
7. Mani I, Gilbert P, Sankale JL, Eisen G, Mboup S, Kanki PJ. Intrapatient diversity and its correlation with viral setpoint in human immunodeficiency virus type 1 CRF02_A/G-IbNG infection. J Virol 2002;76:10745–10755. [PubMed: 12368317]
8. Ripamonti C, Leitner T, Lauren A, Karlsson I, Pastore A, Cavarelli M, et al. Biological and genetic evolution of HIV type 1 in two siblings with different patterns of disease progression. AIDS Res Hum Retroviruses 2007;23:1531–1540. [PubMed: 18160011]
9. Borrow P, Lewicki H, Wei X, Horwitz M, Peffer N, Meyers H, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape variants. Nat Med 1997;3:205–211. [PubMed: 9018240]
10. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 1997;3:212–217. [PubMed: 9018241]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

11. Barouch DH, Kunstman J, Kuroda MJ, Schmitz JE, Santra S, Peyerl FW, et al. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. Nature 2002;415:335–339. [PubMed: 11797012]
12. Koenig S, Conley AJ, Brewah YA, Jones GM, Leath S, Boots LJ, et al. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. Nat Med 1995;1:330–336. [PubMed: 7585062]
13. Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, et al. Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nat Med 2004;10:275–281. [PubMed: 14966520]
14. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, et al. HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 2004;10:282–289. [PubMed: 14770175]
15. Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, Reddy S, et al. 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 2007;81:8346–8351. [PubMed: 17507468]
16. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, et al. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog 2008;4:e1000033. [PubMed: 18369479]
17. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J Virol 2006;80:3617–3623. [PubMed: 16537629]
18. Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 2003;100:4144–4149. [PubMed: 12644702]
19. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. Nature 2003;422:307–312. [PubMed: 12646921]
20. Deeks SG, Schweighardt B, Wrin T, Galovich J, Hoh R, Sinclair E, et al. Neutralizing antibody responses against autologous and heterologous viruses in acute versus chronic human immunodeficiency virus (HIV) infection: evidence for a constraint on the ability of HIV to completely evade neutralizing antibody responses. J Virol 2006;80:6155–6164. [PubMed: 16731954]
21. Geffin R, Hutto C, Andrew C, Scott GB. A longitudinal assessment of autologous neutralizing antibodies in children perinatally infected with human immunodeficiency virus type 1. Virology 2003;310:207–215. [PubMed: 12781708]
22. Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, Paxinos E, et al. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci U S A 2005;102:18514–1851. [PubMed: 16339909]
23. Martin HL, Nyange PM, Richardson BA, Lavreys L, Mandaliya K, Jackson DJ, et al. Hormonal contraception, sexually transmitted diseases, and the risk of heterosexual transmission of HIV-1. J Infect Dis 1998;178:1053–1059. [PubMed: 9806034]
24. Lavreys L, Baeten JM, Chohan V, McClelland RS, Hassan WM, Richardson BA, et al. Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. Clin Infect Dis 2006;42:1333–1339. [PubMed: 16586394]
25. Sagar M, Lavreys L, Baeten JM, Richardson BA, Mandaliya K, Chohan BH, et al. Infection with multiple HIV-1 variants is associated with faster disease progression. J Virol 2003;77:12921–12926. [PubMed: 14610215]
26. Piantadosi A, Chohan B, Chohan V, McClelland RS, Overbaugh J. Chronic HIV-1 infection frequently fails to protect against superinfection. PLoS Pathog 2007;3:e177. [PubMed: 18020705]
27. Rainwater S, Devange S, Sagar M, Ndinya-Achola J, Mandaliya K, Kreiss JK, Overbaugh J. No evidence for rapid subtype C spread within an epidemic in which multiple subtypes and intersubtype recombinants circulate. AIDS Res Hum Retroviruses 2005;21:1060–1065. [PubMed: 16379611]
28. Sagar M, Kirkegaard E, Lavreys L, Overbaugh J. Diversity in HIV-1 envelope V1-V3 sequences early in infection reflects sequence diversity throughout the HIV-1 genome but does not predict the extent of sequence diversity during chronic infection. AIDS Res Hum Retroviruses 2006;22:430–437. [PubMed: 16706620]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

29. Rousseau CM, Nduati RW, Richardson BA, John-Stewart GC, Mbori-Ngacha DA, Kreiss JK, Overbaugh J. Association of levels of HIV-1-infected breast milk cells and risk of mother-to-child transmission. J Infect Dis 2004;190:1880–1888. [PubMed: 15499546]
30. Benki S, McClelland RS, Emery S, Baeten JM, Richardson BA, Lavreys L, et al. Quantification of genital human immunodeficiency virus type 1 (HIV-1) DNA in specimens from women with low plasma HIV-1 RNA levels typical of HIV-1 nontransmitters. J Clin Microbiol 2006;44:4357–4362. [PubMed: 17050820]
31. Maddison, D.; Maddison, W. Macclade 4: analysis of phylogeny and character evolution. Sunderland, MA: Sinauer; 2005.
32. Zwickl, DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Austin, Texas: The University of Texas; 2006.
33. Swofford, D. PAUP: phylogenetic analysis using parsimony EDN. Vol. 3. Champaign: Illinois Natural History Survey; 1991.
34. Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. Bioinformatics 2000;16:400–401. [PubMed: 10869039]
35. Long EM, Rainwater SMJ, Lavreys L, Mandaliya K, Overbaugh J. HIV type 1 variants transmitted to women in Kenya require the CCR5 coreceptor for entry, regardless of the genetic complexity of the infecting virus. AIDS Res Hum Retroviruses 2002;18:567–576. [PubMed: 12036486]
36. Blish C, Nedellec R, Mandaliya K, Mosier D, Overbaugh J. HIV-1 subtype A envelope variants from early in infection have variable sensitivity to neutralization and to inhibitors of viral entry. AIDS 2007;6:693–702. [PubMed: 17413690]
37. Wu X, Parast AB, Richardson BA, Nduati R, John-Stewart G, Mbori-Ngacha D, et al. Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. J Virol 2006;80:835–844. [PubMed: 16378985]
38. Stata Statistical Software: Release 9. Vol. 9. College Station, Texas: StataCorp; 2005.
39. Luo M, Blanchard J, Maclean I, Brunham R. Identification of a novel DRB1 allele, DRB1*1112, by sequence-based DRB typing. Tissue Antigens 2001;57:85–86. [PubMed: 11169266]
40. Peters HO, Mendoza MG, Capina RE, Luo M, Mao X, Gubbins M, et al. An integrative bioinformatic approach for studying escape mutations in human immunodeficiency virus type 1 gag in the Pumwani sex worker cohort. J Virol 2008;82:1980–1992. [PubMed: 18057233]
41. Bird TG, Kaul R, Rostron T, Kimani J, Embree J, Dunn PP, et al. HLA typing in a Kenyan cohort identifies novel class I alleles that restrict cytotoxic T-cell responses to local HIV-1 clades. AIDS 2002;16:1899–1904. [PubMed: 12351949]
42. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol 2007;3:e29. [PubMed: 17305421]
43. Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, Nyange P, et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol 1995;69:8118–8122. [PubMed: 7494333]
44. Long EM, Martin HL Jr, Kreiss JK, Rainwater SM, Lavreys L, Jackson DJ, et al. Gender differences in HIV-1 diversity at time of infection. Nat Med 2000;6:71–75. [PubMed: 10613827]
45. Brown BK, Wieczorek L, Sanders-Buell E, Rosa Borges A, Robb ML, Birx DL, et al. Cross-clade neutralization patterns among HIV-1 strains from the six major clades of the pandemic evaluated and compared in two different models. Virology 2008;375:529–538. [PubMed: 18433824]
46. Polonis VR, Brown BK, Rosa Borges A, Zolla-Pazner S, Dimitrov DS, Zhang MY, et al. Recent advances in the characterization of HIV-1 neutralization assays for standardized evaluation of the antibody response to infection and vaccination. Virology 2008;375:315–320. [PubMed: 18367229]
47. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, et al. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. AIDS 2008;22:1277–1286. [PubMed: 18580606]
48. Geels MJ, Cornelissen M, Schuitemaker H, Anderson K, Kwa D, Maas J, et al. Identification of sequential viral escape mutants associated with altered T-cell responses in a human immunodeficiency virus type 1-infected individual. J Virol 2003;77:12430–12440. [PubMed: 14610167]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

49. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan KM, et al. 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 2005;79:13239–13249. [PubMed: 16227247]

50. Patterson S, English NR, Longhurst H, Jain P, Balfe P, Knight SC. Relationship between V3 sequences of HIV type 1 from plasma, T cells, and dendritic cells suggests that plasma virus may arise from different cell sources. AIDS Res Hum Retroviruses 2001;17:179–184. [PubMed: 11177397]

51. Voronin Y, Chohan B, Emerman M, Overbaugh J. Primary isolates of human immunodeficiency virus type 1 are usually dominated by the major variants found in blood. J Virol 2007;81:10232–10241. [PubMed: 17652386]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

		gag		*env*	
			p=0.002		
Early infection diversity		**0.2 (0-1.3)**	↔	**0.3 (0-6.4)**	
	p=0.007 ↕		p<0.001		↕ p=0.008
Chronic infection diversity		**1.1 (0.5-2.6)**	↔	**2.7 (0.9-7.8)**	
	p<0.001 ↕		p=0.001		↕ p<0.001
Chronic infection divergence		**1.7 (0.4-3.1)**	↔	**3.1 (1.1-8.5)**	

Fig. 1. Correlations between evolution in *gag* and *env* during early and chronic infection

The median (and range) percentage diversity and divergence in each gene (*gag* and *env*) in early and chronic infection is shown for the 37 women in this study. Lines with horizontal arrows and *P* values indicate the results of comparisons of diversity and divergence between genes (Spearman's correlation test). Within each gene, lines with vertical arrows and *P* values indicate the results of comparisons of diversity between time points and comparisons of diversity and divergence at the chronic infection time point (Spearman's correlation test). In all tests, rho was at least 0.44.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

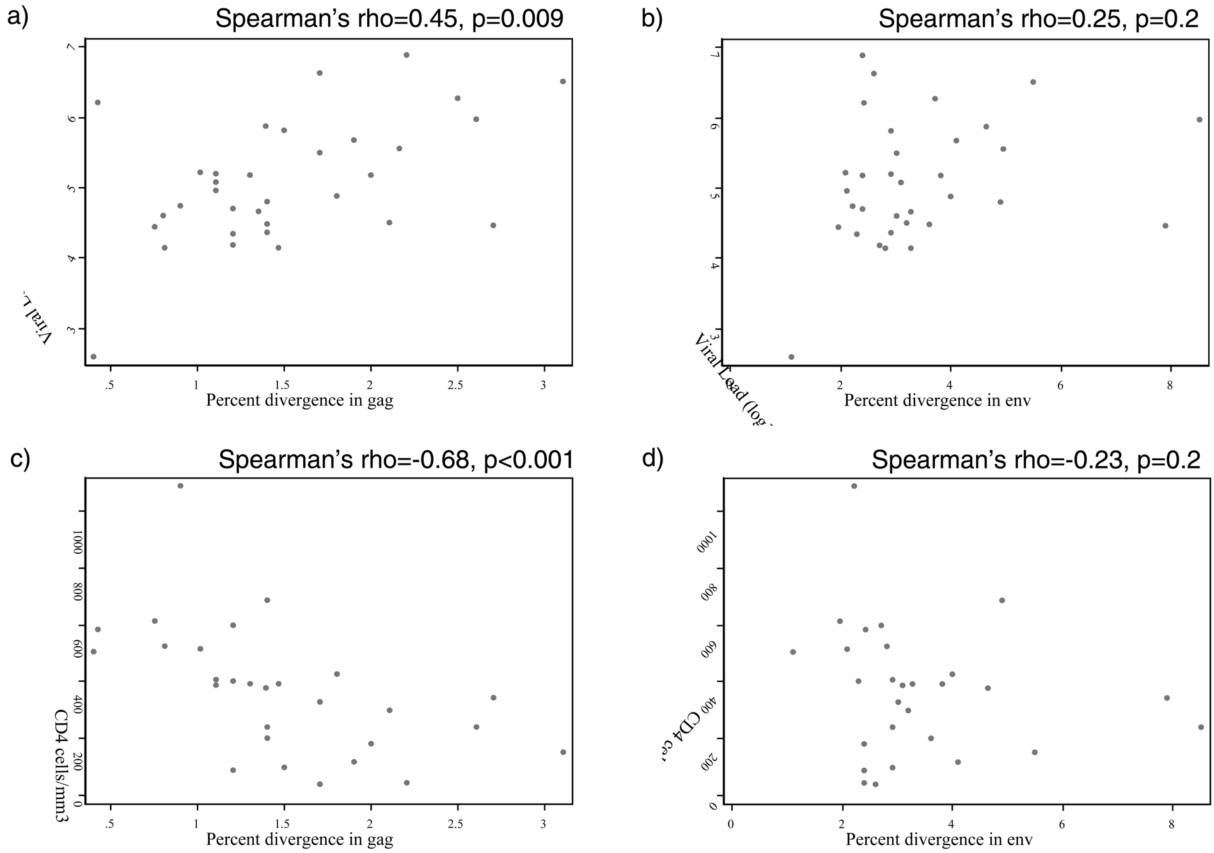

Fig. 2. Associations between *gag* and *env* divergence and plasma viral load and CD4 cell count during chronic infection

Plasma viral load at the chronic infection time point (available for 33/37 individuals) is plotted against each individual's median percentage divergence in *gag* (a) and *env* (b). CD4 cell count measured within 6 months of the chronic infection time point (available for 28/37 individuals) is plotted against each individual's median percentage divergence in *gag* (c) and *env* (d). Each point represents the viral load and mean percentage divergence for sequences from one individual. Comparisons were performed using the Spearman's correlation test, and results are shown at the top of each graph.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

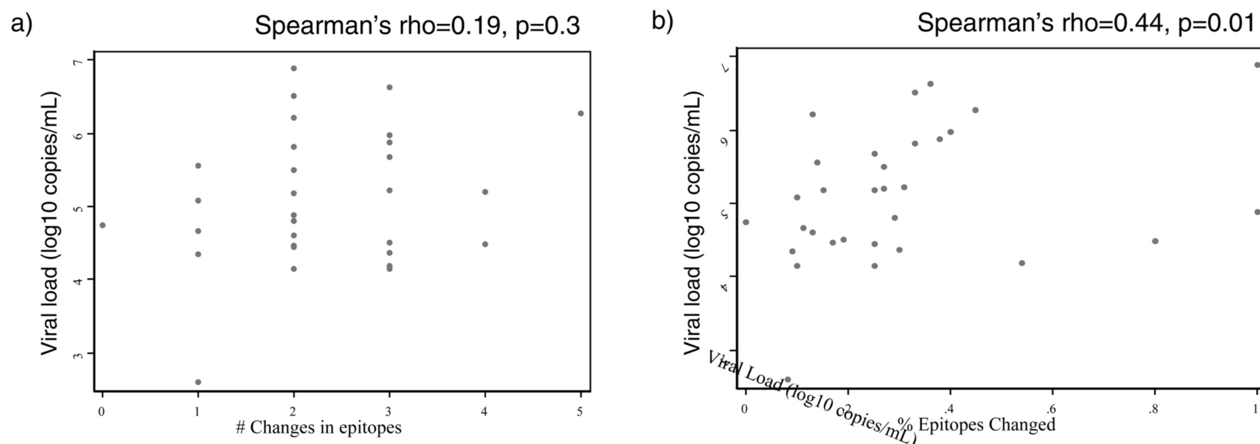

Fig. 3. Associations between viral load during chronic infection and the number and percentage of *gag* cytotoxic T lymphocyte epitopes changed

Plasma viral load measured at the chronic infection time point is plotted against the number of amino acid changes in *gag* CTL epitopes (a) and against the percentage of total published *gag* CTL epitopes that contain at least one amino acid change (b). Each point represents one individual. Comparisons were performed using the Spearman's correlation test, and results are shown at the top of each graph. CTL, cytotoxic T lymphocyte.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Plasma Subject ID	Virus: Subtype B SF162	Subtype A Q461D1	Q168B23	Q769H5	Q842D16	Q259D2.17	NAb Breadth Score
A028	50	619	2858	1768	50	50	3
A039	50	1688	1181	122	4000	112	5
A078	50	697	1915	440	50	50	3
A210	2209	324	154	50	50	50	1
A255	565	2761	2172	831	112	113	6
A261	689	4000	2028	1651	50	50	4
A268	50	50	50	50	50	50	0
A281	50	156	50	50	50	50	0
A284	340	215	592	125	50	50	3
A520	2031	1691	4000	151	50	50	4
A584	174	230	799	130	109	139	4
A692	4000	1197	458	541	50	50	4
A779	415	100	192	50	50	50	1
A966	4000	872	791	288	50	1454	5
B368	429	6653	694	50	50	50	3
B374	50	4000	552	460	50	50	3
B670	1071	332	158	50	50	50	1
C036	506	929	844	401	50	213	5
C152	569	1235	1959	804	142	265	6
C322	197	327	310	170	129	202	3
C511	1124	507	480	281	184	334	6
C805	210	579	560	50	50	50	2
C890	246	50	188	50	167	112	2
D385	50	112	219	50	50	50	0
Assay median	**249**	**373**	**311**	**120**	**50**	**50**	

Fig. 4. Example of neutralization assay with plasma from a representative subset of 24 women in this study

This table lists the IC_{50}, the reciprocal dilution of plasma that neutralizes 50% of virus, for each plasma sample (rows) screened against each of six panel viruses (columns). The highest plasma dilution tested was 1: 100, and samples that were not neutralized at this dilution were given the reciprocal IC_{50} value of 50. IC_{50} values that are higher than the assay median for each virus are highlighted in gray. For each plasma sample, the neutralizing antibody breadth score was calculated as the number of panel viruses (out of six) neutralized at a IC_{50} higher than the assay median. This assay was repeated in triplicate for each plasma sample from 36 out of 37 individuals, and the average of three scores was used in later analyses. IC_{50}, 50% maximal inhibitory concentration; NAb, neutralizing antibody.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Summary of *gag* and *env* evolution among 37 individuals.

	gag **Median (range)**	***env*** **Median (range)**	**Wilcoxon** ***P***
Synonymous divergence (%)[a]	2.4 (0.5–6.7)	2.1 (0–11.1)	0.11
Nonsynonymous divergence (%)[a]	1.0 (0.3–2.2)	3.1 (0–6.8)	<0.001
dN/dS[b]	0.4 (0.2–1.0)	1.5 (0.3–2.7)	<0.001
Amino acid changes (%)[c]	3.4 (1.0–6.8)	8.7 (2.4–15.8)	<0.001

dN/dS, ratio of nonsynonymous to synonymous divergence; MRCA, most recent common ancestor.

[a]For each individual, the percentage synonymous divergence and nonsynonymous divergence were calculated as the mean synonymous or nonsynonymous distance from a chronic infection sequence to the MRCA of the early infection sequences.

[b]For each individual, dN/dS was calculated as the mean dN/dS value comparing each chronic infection sequence to the MRCA of the early infection sequences.

[c]For each individual, the percentage amino acid changes was calculated as the percentage of amino acid positions at which at least two chronic infection sequences shared an amino acid that was different from the MRCA of the early infection sequences.