OPEN ACCESS Freely available online

PLoS PATHOGENS

HIV-1 Envelope Subregion Length Variation during Disease Progression

Marcel E. Curlin[1]*, Rafael Zioni[2¤], Stephen E. Hawes[3], Yi Liu[4], Wenjie Deng[4], Geoffrey S. Gottlieb[1], Tuofu Zhu[2,4], James I. Mullins[1,2,4]

1 Department of Medicine, University of Washington School of Medicine, Seattle, Washington, United States of America, **2** Department of Laboratory Medicine, University of Washington School of Medicine, Seattle, Washington, United States of America, **3** Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, United States of America, **4** Department of Microbiology, University of Washington School of Medicine, Seattle, Washington, United States of America

Abstract

The V3 loop of the HIV-1 Env protein is the primary determinant of viral coreceptor usage, whereas the V1V2 loop region is thought to influence coreceptor binding and participate in shielding of neutralization-sensitive regions of the Env glycoprotein gp120 from antibody responses. The functional properties and antigenicity of V1V2 are influenced by changes in amino acid sequence, sequence length and patterns of N-linked glycosylation. However, how these polymorphisms relate to HIV pathogenesis is not fully understood. We examined 5185 HIV-1 gp120 nucleotide sequence fragments and clinical data from 154 individuals (152 were infected with HIV-1 Subtype B). Sequences were aligned, translated, manually edited and separated into V1V2, C2, V3, C3, V4, C4 and V5 subregions. V1-V5 and subregion lengths were calculated, and potential N-linked glycosylation sites (PNLGS) counted. Loop lengths and PNLGS were examined as a function of time since infection, CD4 count, viral load, and calendar year in cross-sectional and longitudinal analyses. V1V2 length and PNLGS increased significantly through chronic infection before declining in late-stage infection. In cross-sectional analyses, V1V2 length also increased by calendar year between 1984 and 2004 in subjects with early and mid-stage illness. Our observations suggest that there is little selection for loop length at the time of transmission; following infection, HIV-1 adapts to host immune responses through increased V1V2 length and/or addition of carbohydrate moieties at N-linked glycosylation sites. V1V2 shortening during early and late-stage infection may reflect ineffective host immunity. Transmission from donors with chronic illness may have caused the modest increase in V1V2 length observed during the course of the pandemic.

Citation: Curlin ME, Zioni R, Hawes SE, Liu Y, Deng W, et al. (2010) HIV-1 Envelope Subregion Length Variation during Disease Progression. PLoS Pathog 6(12): e1001228. doi:10.1371/journal.ppat.1001228

Editor: Alexandra Trkola, University of Zurich, Switzerland

Received May 26, 2010; **Accepted** November 11, 2010; **Published** December 16, 2010

Funding: This work was supported by NIH (http://www.nih.gov/) grants AI52791, AI047734, AI058894, AI57005, AI49109, AI45402, AI55336 and the Computational Biology Core of the University of Washington Center for AIDS Research (http://depts.washington.edu/cfas/) (AI27757). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cemarcel@u.washington.edu

¤ Current address: Stanford University Division of Infectious Diseases, Stanford University Medical Center, Stanford, California, United States of America

Introduction

The gp120 portion of the HIV-1 envelope protein (Env) mediates attachment prior to fusion with the host cell membrane during target cell infection. gp120 has five hypervariable regions (V1–V5) bounded by cysteine residues and separated by four relatively "constant" regions (C1–C4) [1–3]. Gp120 is notable for its sequence variation, which may arise through recombination and point mutation, as well as by insertion and deletion of one or more nucleotides. Insertion and deletion events (indels) occur throughout *env* but are maintained through positive selection particularly within the hypervariable loops, which thereby may acquire significant length variation [4], The third hypervariable region is known to encode the primary determinants of coreceptor usage specificity [5–7], as well as epitopes recognized by humoral [8,9] and cellular [10,11] immune responses. V3 loop sequence variation has been extensively studied, and correlated with changes in host cell range, cytopathogenicity, and disease progression [12–14].

The V1V2 region in particular is characterized by a high degree of length polymorphism, sequence variation, and predicted N-linked glycosylation sites (PNLGS) [15–20], each of which may affect viral attachment, coreceptor usage and recognition by neutralizing antibodies [20,21]. Comparison of structural models of gp120 and gp120 bound to CD4 and a chemokine coreceptor have yielded considerable insight into the functional roles played by V1V2 and V3 during viral attachment [22,23]. In the unbound gp120 conformation, the V2 loop partially obscures V3 and other gp120 residues involved in coreceptor binding. Binding to CD4 induces conformational changes that expose the coreceptor binding site on gp120, including residues from V1V2, V3 and other regions [22,24].

Numerous studies have suggested that sequence variation in V1V2 influences host cell range and/or syncytium-inducing (SI) phenotype [25–31]. For example, Toohey demonstrated that recombinant chimeric clones with a V1V2 region from macrophage-tropic HIV-1 strains replicated efficiently in macrophages, whereas clones with the V1V2 region from lymphotropic strains did not [31]. However, not all studies have been concordant on the role of V1V2 in viral replication kinetics, cell range and transmission [15–19,32]. For example, Pastore showed that

Author Summary

The HIV envelope gene (*env*) encodes viral surface proteins (Env) that are vital to the basic processes used by the virus to infect and cause disease in humans. Adaptations in *env* determine which cells the virus can infect, and permit the virus to avoid elimination by the immune system. *Env* is one of the most variable genes known, and it can change dramatically over time in a single individual. However, Env-host cell interactions are complex and incompletely understood, and changes in this viral protein during infection have not yet been systematically described. We examined a large number of *env* sequences from 154 individuals at various stages of HIV infection but who had never received antiretroviral treatment. We found that the *env* V1V2 region lengthens during chronic infection and becomes more heavily glycosylated. However, these changes partially reverse during late-stage illness, possibly in response to a weakening host immune system. V1V2 lengths are also increasing over time in the epidemic at large, possibly related to the epidemiology of HIV transmission within the subtype B epidemic. These results provide fundamental insights into the biology of HIV.

sequence changes in V1V2 could rescue otherwise lethal mutations in V3 associated with a change in coreceptor usage [33], and V2 polymorphisms have also been linked with restriction to CCR5 coreceptor usage [16]. In contrast, Wang et al found no relationship between SI phenotype and V1V2 sequence, length, distribution of PNLGS or charge [32].

The V1V2 region also appears to be an important determinant of sensitivity to neutralizing antibodies [34–38]. The V1V2 region evolves under positive natural selection *in vivo* [4,39–41], and an inverse relationship between V1–V4 length and neutralization susceptibility has been demonstrated in subtypes A [20], B [34–38] and C [42]. Tellingly, laboratory strains lacking V1V2 may still replicate efficiently *in vitro*, but appear to be especially sensitive to antibody neutralization [43,44]. Consistent with this observation, viral strains with shorter and less glycosylated V1V4 regions have been reported to preferentially replicate in subjects newly infected with HIV-1 subtype C [45] (where presumably an effective neutralizing antibody response has not had time to emerge), and similar observations have been made concerning the V1V2 loop in individuals recently infected by HIV-1 subtype A [19]. However, we and others have not observed this effect in HIV-1 subtype B [19,46,47].

Despite these reports, the relationship between V1V2 region length polymorphism and disease progression remains unclear. In two small longitudinal studies, elongation of V1 and V2 was noted in long-term nonprogressors (LTNP), but not within individuals progressing rapidly to AIDS [15–19]. In a third study, no clear relationship between V1V2 length variation and disease progression was observed [48]. Lastly, some investigators postulate that V1V2 length changes positively correlate with the pace of disease progression [16,19], while others have suggested that V1V2 length increase may be a correlate of delayed progression to AIDS [18].

Thus, our understanding of the role of the V1V2 loop in influencing HIV pathogenesis remains incomplete and is challenged by several contradictory observations. To more fully characterize HIV envelope subregion variability and to clarify the associations between subregion length variation, glycosylation, and disease progression, we have comprehensively examined length and glycosylation of each gp120 subregion as a function of clinical parameters in a large collection HIV-1 subtype B infected individuals.

Methods

Ethics statement

This study was performed using publicly available data from the Los Alamos database, and previously unpublished experimental data obtained at the University of Washington. Unpublished data were obtained and analyzed with written informed consent of study participants, and approval by the University of Washington Institutional Review Board.

Patient selection

We analyzed new and published HIV-1 envelope gene sequences and associated clinical data from all available subjects in the Seattle Primary Infection Cohort (PIC) [49], the Multicenter AIDS Cohort Study (MACS) [50], and from the Los Alamos National Laboratories HIV database (HIVDB) (http://www.hiv.lanl.gov/content/hiv-db/mainpage.html) not meeting pre-specified exclusion criteria. Subjects were excluded from this study if younger than 18 years of age or if there was any history of antiretroviral therapy prior to sampling as determined by patient report and clinical records (MACS, PIC) or as indicated in the methods section of published reports (HIVDB), unless otherwise noted. All subjects considered in the cross-sectional and longitudinal analyses were infected with HIV-1 subtype B, except for two subjects infected with HIV-1 subtype A who were included in longitudinal analyses, but were excluded from cross-sectional analyses. (Additional subtypes were considered in analyses of *env* subregion length change during transmission, presented in Text S1, Section 8). Clinical data retrieved included CD4 count, viral load, time since infection, and treatment history. Sequence data were only accepted if directly derived from plasma or PBMC without an intervening step involving viral propagation *in vitro*. In some cases, individual authors were consulted to resolve clinical or methodological ambiguities. Accession numbers for published sequences are provided in Table S1. Gene sequence data used in this study are available at http://mullinslab.microbiol.washington.edu/publications/curlin_2010/.

Subject groups

Viral gene sequence data were considered in both cross-sectional (Table 1) and longitudinal analyses (Table 2). The cross-sectional dataset included only plasma and PBMC sequences derived from individuals infected with subtype B (see results, and Table 1). Sequences were triaged by author, database identifier and associated clinical data to exclude duplicate entries. To assess the role of stage of illness on loop length variation, subjects were divided into four non-overlapping groups; *group* C_x1 subjects were sampled within two months of the estimated time of infection. *Group* C_x2 subjects were sampled between two months and three years following infection. *Group* C_x3 subjects were sampled at times >3 years post infection. *Group* C_x4 was comprised of all individuals meeting 1993 CDC criteria for AIDS when sampling occurred (generally CD4 count $<200/mm^3$), regardless of time since infection.

The longitudinal dataset was derived from 20 subjects infected with subtype B and 2 individuals infected with subtype A, from the PIC cohort and from previous reports [18,51–55], in whom data were available from two or more timepoints (see results, and Table 2). All intra-individual longitudinal comparisons were made between sequences obtained from the same compartment (e.g., plasma *vs.* plasma). Individuals partitioned into group *L1* (N = 15)

Table 1. Distribution of subjects, samples and sequences in cross-sectional analyses.

Cross-sectional Data									
Group	**Subjects**	**Samples**	**Sequences**						
			V1V2	**C2**	**V3**	**C3**	**V4**	**C4**	**V5**
TOTAL	152	453	1922	1275	4407	3616	4406	4405	4407
MACS	27	227	682	682	2567	2568	2567	2569	2569
PIC	43	78	541	390	846	845	845	845	847
LANL	82	148	699	203	994	203	994	991	991
Sequences by category									
PBMC	62	225	385	65	2193	1675	2193	2193	2193
Plasma	90	228	1537	1210	2214	1941	2213	2212	2214
Asia	11	16	176	0	0	0	0	0	0
North America	111	362	1585	1207	3725	3548	3724	3726	3728
South America	5	5	5	5	5	5	5	5	5
Western Europe	25	70	156	63	677	63	677	674	674
Stage 1	41	42	418	365	383	383	383	384	384
Stage 2	40	146	872	741	1540	1483	1539	1540	1542
Stage 3	22	156	220	94	1586	1029	1586	1583	1583
Stage 4	11	63	170	35	631	631	631	631	631
Unknown Stage	38	46	242	40	267	90	267	267	267

Number of subjects, samples, and available coding sequences by cohort, gene region, anatomical site, geographic location, and stage of illness.
doi:10.1371/journal.ppat.1001228.t001

Table 2. Distribution of subjects, samples and sequences in longitudinal analyses.

Longitudinal Data									
Group	**Subjects**	**Samples**	**Sequences**						
			V1V2	**C2**	**V3**	**C3**	**V4**	**C4**	**V5**
TOTAL	22	83	807	30	155	155	155	155	155
PIC	1	15	180	30	155	155	155	155	155
LANL	21	68	627	0	0	0	0	0	0
Sequences by category									
Culture	NA	1	33	0	0	0	0	0	0
Cervical Swab	NA	7	37	0	0	0	0	0	0
PBMC	N/A	43	397	0	0	0	0	0	0
Plasma	N/A	32	340	30	155	155	155	155	155
Subtype A	2	26	172	0	0	0	0	0	0
Subtype B	20	57	635	30	155	155	155	155	155
Asia	4	9	102	0	0	0	0	0	0
East Africa	2	26	172	0	0	0	0	0	0
North America	12	40	440	30	155	155	155	155	155
Western Europe	4	8	93	0	0	0	0	0	0
Stage 1	NA	5	63	10	23	23	23	23	23
Stage 2	NA	12	140	10	111	111	111	111	111
Stage 3	NA	10	111	10	21	21	21	21	21
Stage 4	NA	9	86	0	0	0	0	0	0
Unknown Stage	10	47	407	0	0	0	0	0	0

Number of subjects, samples, and available coding sequences by cohort, gene region, anatomical site, HIV subtype, geographic location, and stage of illness.
doi:10.1371/journal.ppat.1001228.t002

did not meet criteria for AIDS at any time prior to the final sample (median follow-up 3.25 years, range 1 to 20.8 years), whereas subjects in group *L2* (N = 7) were reported to have an AIDS-defining illness or peripheral CD4 count $<200/mm^3$ between the first and second samples (median follow-up 2.75 years, range 2 to 4 years).

Nucleic acid isolation, cloning and sequencing

Sequences from the PIC and MACS cohorts (Tables 1 & 2) were obtained from plasma or PBMC by standard methods [56,57], using safeguards to prevent contamination and template resampling [58]. Briefly, PCR amplification was performed using Taq polymerase (Bioline) with primers ED3 and BH2 [59] (first round) followed by ED5 and DR7 (second round) [60]. PCR products were cloned into a TA TOPO vector (Invitrogen) and selected colonies sequenced under contract using Big Dye dye-terminator protocols. Genbank accession numbers pending submission.

Sequence analysis

Deduced amino acid sequences were aligned using ClustalW [61] and divided into seven subregions; V1V2 (HXB2 nucleotide positions 6615–6812), C2 (HXB2 6813-7109), V3 (HXB2 7110–7217), C3 (HXB2 7218–7376), V4 (HXB2 7377–7478), C4 (HXB2 7479–7556), and V5 (HXB2 7557–7637). Alignments were manually edited and subregion lengths were counted using MacClade. PNLGS were counted using NetNGlyc.1 (http://www.cbs.dtu.dk/services/NetNGlyc/). Coreceptor usage (CCR5 *vs.* CXCR4 tropism) was predicted for all available subtype B V3 loop sequences, using the Position-Specific Substitution Method (PSSM) [62], Geno2pheno [63] and two other machine learning algorithms [64,65] (hereafter denoted PSSM, G2P, PGRC and BMLC, respectively). For G2P coreceptor usage predictions, we selected the standard 10% false positivity threshold, and PGRC predictions were based on the support vector machine (SVR) user option. Estimated time since infection was calculated for all data entries. When time was reported as time since onset of symptoms or time post seroconversion (SC), symptoms and seroconversion were assumed to occur at 14 days and 42 days after infection, respectively [66,67]. Date of seroconversion was assumed to occur at the midpoint between most recent negative serological test and first reported positive test, unless additional information was available.

Statistical analysis

For cross-sectional analyses, univariate and multivariate regressions were conducted assessing subregion lengths and number of glycosylation sites as a function of time since infection, stage of disease, CD4 count, HIV viral load, adjusting for sample source (plasma *vs.* PBMC), and date of sampling (calendar year). In regression analyses, to allow direct comparisons of the effect of each variable on V1V2 length and/or glycosylation, we compared β values (i.e., regression coefficients scaled such that each variable is equivalent to having a mean value of 0 and a standard deviation of 1). Generalized estimating equations (GEE) were utilized to account for non-independence of data points [68–70], and an exchangeable correlation structure was assumed. This method adjusts for the correlation of multiple sequences nested within a sample as well as multiple samples per patient. As an additional means of verifying that analysis outcomes were not influenced by data linkage, regression analyses were performed on replicate data subsets reconstituted from the original data by random resampling, including analyses on 100 data subsets each obtained by using one randomly selected sequence from each individual (See Text S1 section S2). To ensure that results were not unduly influenced by outlying sequences with extremely short or long loop lengths, analyses were repeated after excluding sequences representing the shortest 5% and longest 5% of the V1V2 loops in the dataset. For the longitudinal dataset, multivariate linear regressions were conducted assessing V1V2 length and number of glycosylation sites as a function of time since infection within a person, and the mean rate of change per year was estimated. Statistical analyses were performed using SAS version 9.1 (SAS Institute, Cary, NC).

Results

Sequence data

We obtained 5185 partial length HIV-1 env gene sequences for cross-sectional and longitudinal analysis by the methods described above (Tables 1 & 2). Sequences were isolated from 475 samples obtained from 154 individuals, including 27 from the MACS, 43 from the Seattle PIC and 84 from the HIVDB. Study subjects resided in North America (N = 116), Western Europe (N = 25), East Africa (N = 2), and Asia (N = 11), contributed a median of 14 sequences (range 1–287) and included persons in stages 1 (N = 41), 2 (N = 62), 3 (N = 40), and 4 (N = 27) of infection (note that some subjects contributing to the longitudinal analysis were included at more than one stage of infection). Sequences were derived from plasma (N = 2495), PBMC (N = 2620) and other sites (N = 70). Sequences were of subtype B (N = 5013) and subtype A (N = 172). All subtype A sequences and sequences derived from sites other than blood were excluded from cross-sectional analyses, but were considered as special cases under longitudinal analyses (sequence data available at: *webaddress pending acceptance*).

Cross-sectional analyses

Variation in sequence length and glycosylation. The V1V2, V4 and V5 hypervariable regions displayed heterogeneity in lengths up to approximately 2-fold in the 152 individuals examined. V1V2 was the most variable region, with loop lengths ranging from 50 to 99 amino acids (mean = 68), while V4 and V5 loop lengths ranged from 19 to 44 (mean = 32), and 14 to 36 (mean = 28) amino acids, respectively. In contrast, the V3 loop and the C2, C3 and C4 regions showed relatively little length variation (Figure 1). The subregions with the greatest number of potential glycosylation sites were V1V2 (mean 6 sites, range 0–12), C2 (mean 5, range 3–8) and V4 (mean 5, range 1–7). V3, C3 and V5 were more modestly glycosylated (mean = 1, 3, and 2, respectively, with a maximum of 5 glycosylated sites), whereas C4 rarely contained potential glycosylation sites (1 site was found in 8 of 4403 sequences).

Relationship between V1V2 loop length, sample features and clinical factors – univariate analyses. (And see Text S1, sections S1, S3, S6, and Figures S1, S2 and S12.) We examined V1V2 loop lengths as a function of year of sampling and specimen type (plasma *vs.* PBMC). In separate univariate GEE analyses, V1V2 length increased with calendar year of sampling ($\beta = 1.62$ increase in V1V2 length per year; $p = 0.003$, Figure 2, lower panel) and trended towards greater length in PBMC, though not significantly ($\beta = 1.70$ for PBMC compared to plasma; $p = 0.11$). We then examined individual subregion lengths as a function of time since infection, clinical stage, CD4 counts, and HIV plasma viral load. In separate GEE regression analyses, V1V2 length was significantly correlated with time since infection ($\beta = 1.00$ increase in V1V2 length per year; $p<0.001$, Figure 2, upper panel) and clinical stage, as subjects with stage 3 ($\beta = 6.36$; $p<0.001$) and stage 4 ($\beta = 3.30$; $p = 0.02$), but not stage 2 ($\beta = 0.80$; $p = 0.4$) had

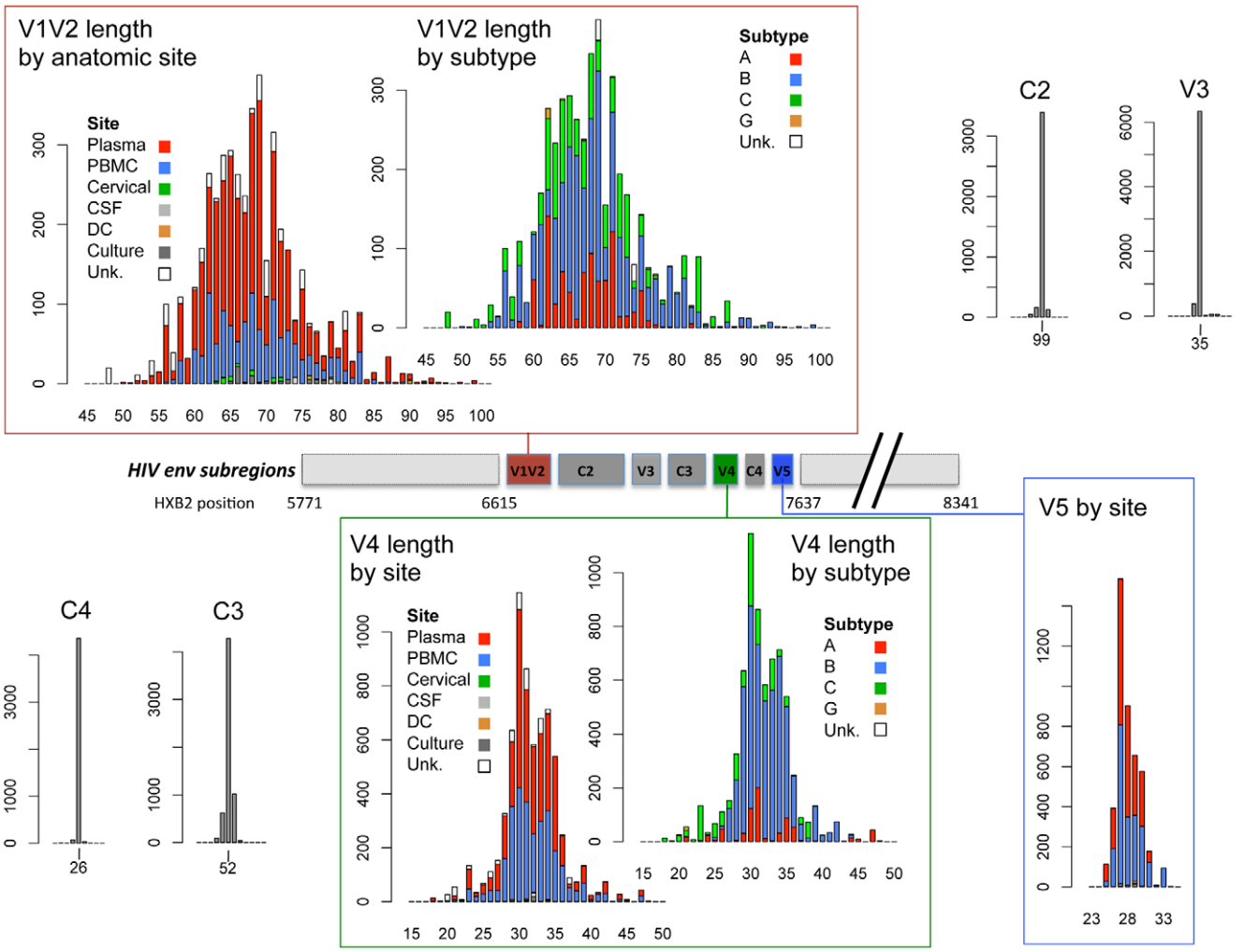


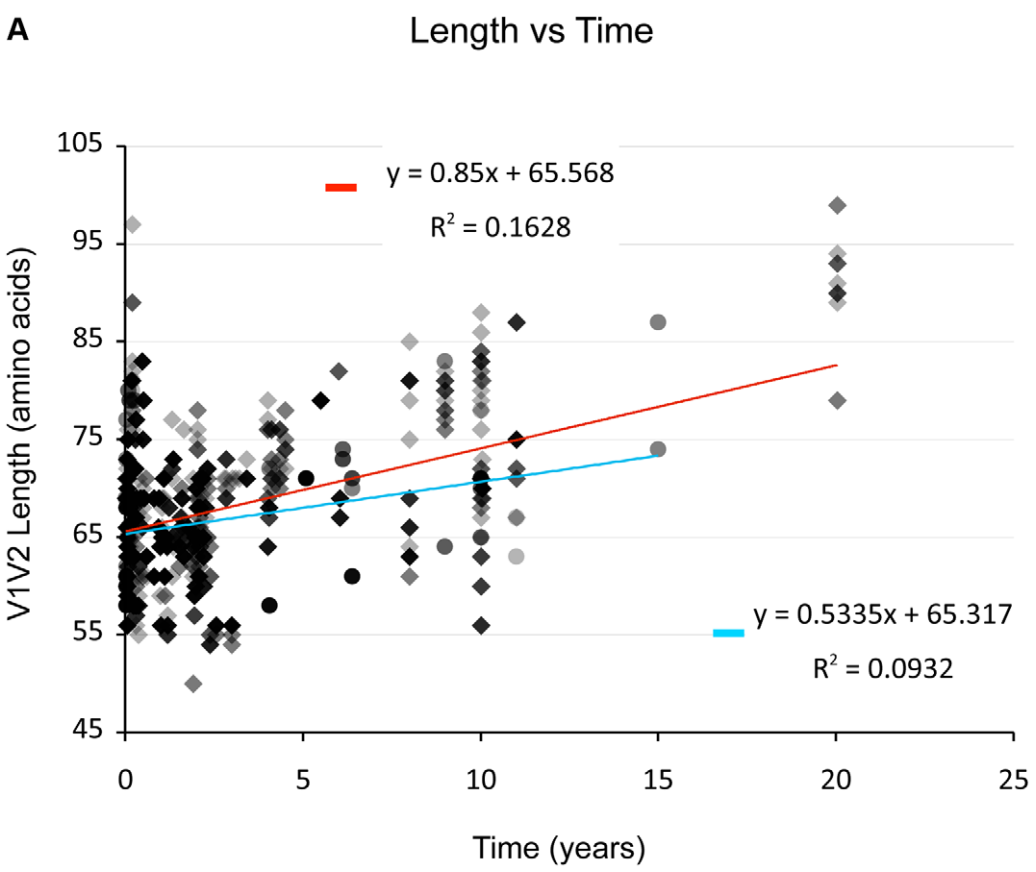
Figure 1. Schematic diagram of HIV-1 *env* subregions (center bar) and distribution of subregion loop lengths (surrounding bar graphs). The center bar depicts the linear arrangement of subregions V1V2 through V5 within the HIV Env gp120 protein. The amino acid length distribution of each subregion is shown in the linked bar graphs, including sequences in the cross-sectional dataset, the longitudinal dataset and the transmission data described in Text S1. Length distributions in V1V2 and V4 data are shown by isolation site (PBMC = blue bars, plasma = red bars, cervical cells = green bars, CSF = light gray bars, dendritic cells = orange bars, cell culture = dark gray bars, and cells from unknown anatomic compartments represented by open bars), and by subtype (subtype B = blue bars, subtype A = red bars, subtype C = green bars, subtype G = orange bars, untyped sequences = open bars). V5 sequences were all of subtype B. X-axis: sequence length (amino acids); Y-axis: number of sequences.

doi:10.1371/journal.ppat.1001228.g001

significantly longer V1V2 lengths compared to subjects with stage 1 infection (Figure 3 and S12). However, V1V2 length did not significantly correlate with either CD4 stratum (<200, 200–500 or >500 cells/ml) or plasma viral load.

Relationship between V1V2 loop length, sample features and clinical factors – multivariate analyses (Table 3). To further understand the interaction between significant variables, we next performed multivariate analyses of V1V2 length vs. time since infection, clinical stage, CD4 level, and HIV viral load after adjusting for calendar year and type of sample. This analysis was performed for all sequences in the dataset, as well as with plasma sequences and PBMC sequences considered separately (Table 3). Overall, V1V2 length was not significantly associated with time since infection, CD4 level, or HIV viral load. However, among sequences derived from plasma, V1V2 length was significantly associated with increased time since infection ($\beta = 0.77$ per year; $p<0.001$). Conversely, among the PBMC sequences, V1V2 length was associated with decreased CD4 counts ($\beta = 8.13$ for CD4 counts between 200 and 500 and $\beta = 6.77$ for CD4 counts less than 200 compared to >500) although the association with the lowest CD4 count group did not reach statistical significance ($p = 0.09$). Among subjects without AIDS (Stages 1 through 3), V1V2 length was associated with time since infection ($\beta = 0.70$ increase in V1V2 length per year; $p<0.001$), even after adjustment for calendar year and type of sample (data not shown). Overall, after adjusting for calendar year and sample type, V1V2 length remained significantly associated with clinical stage, as subjects with stage 3 ($\beta = 6.25$; $p<0.001$) and stage 4 ($\beta = 3.54$; $p = 0.02$), but not stage 2 ($\beta = 0.09$; $p = 0.9$) had significantly longer V1V2 lengths compared to subjects with stage 1 infection. However, V1V2 lengths in subjects with clinical stage 4 were significantly shorter than V1V2 lengths from subjects in stage 3 ($p<0.001$). The findings of increased V1V2 length in stage 3 and 4 infection compared to stage 1 and 2 were similarly noted both among

A


Length vs Time
y = 0.85x + 65.568
R² = 0.1628
y = 0.5335x + 65.317
R² = 0.0932
V1V2 Length (amino acids)
Time (years)
105
95
85
75
65
55
45
0
5
10
15
20
25

B

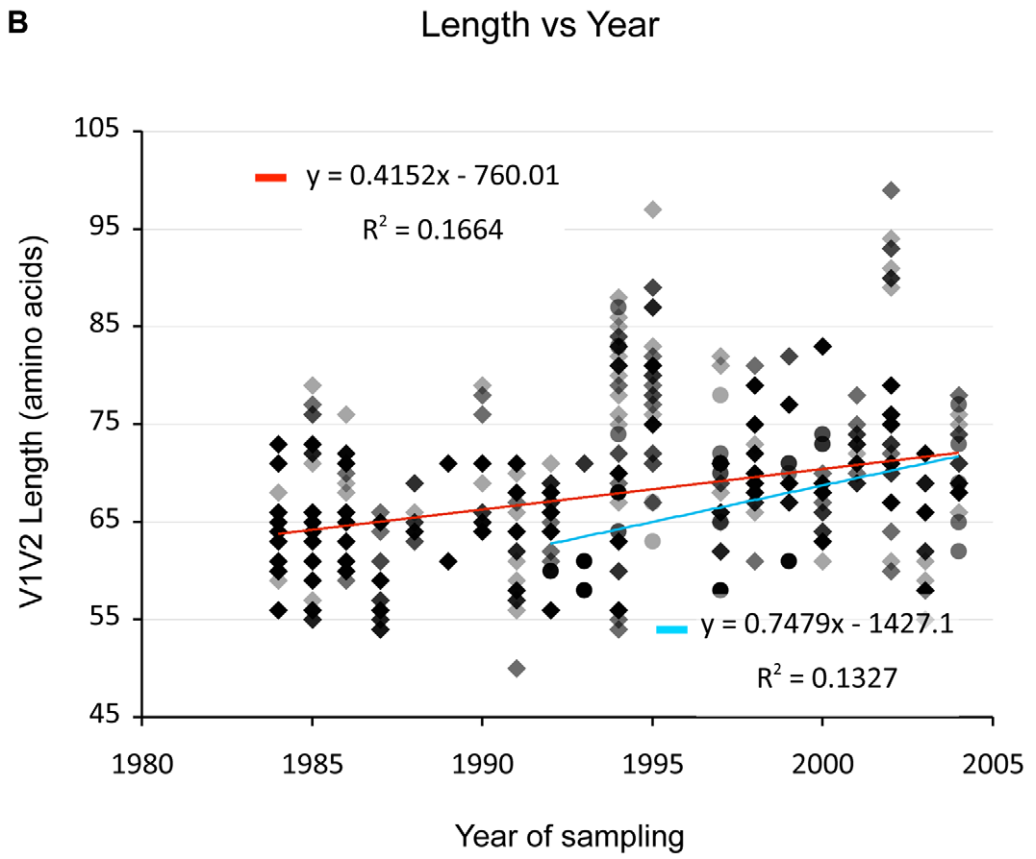

Length vs Year
y = 0.4152x - 760.01
R² = 0.1664
y = 0.7479x - 1427.1
R² = 0.1327
V1V2 Length (amino acids)
Year of sampling
105
95
85
75
65
55
45
1980
1985
1990
1995
2000
2005

Figure 2. V1V2 length *vs.* time since infection (upper panel) and vs. year of sampling (lower panel). Lengths are indicated in amino acids. Overlapping data points appear as darker symbols. Sequences from plasma are represented by diamonds and sequences from PBMC are represented by circles. Regression coefficients and coefficients of determination are shown for univariate linear regression, for plasma (red line) and PBMC (blue line).
doi:10.1371/journal.ppat.1001228.g002

sequences derived from plasma as well as PBMC, although the plasma associations did not reach statistical significance in all cases. In order to assess the potential that the results regarding clinical and viral factors associated with V1V2 length could be driven by unusually short or long sequences, we repeated the above analyses excluding the shortest and longest 5% of V1V2 lengths. Since model coefficients and p-values were similar in this restricted analysis (Table S2), our findings do not appear to be unduly influenced by a small number of outlying small or large sequences (Also see Text S1, section S3 and Figure S6).

As an alternative means of accounting for the variable number of sequences contributed by study subjects, the data was subjected to a resampling analysis, in which each subject contributed a single randomly selected sequence. This process was repeated 100 times,

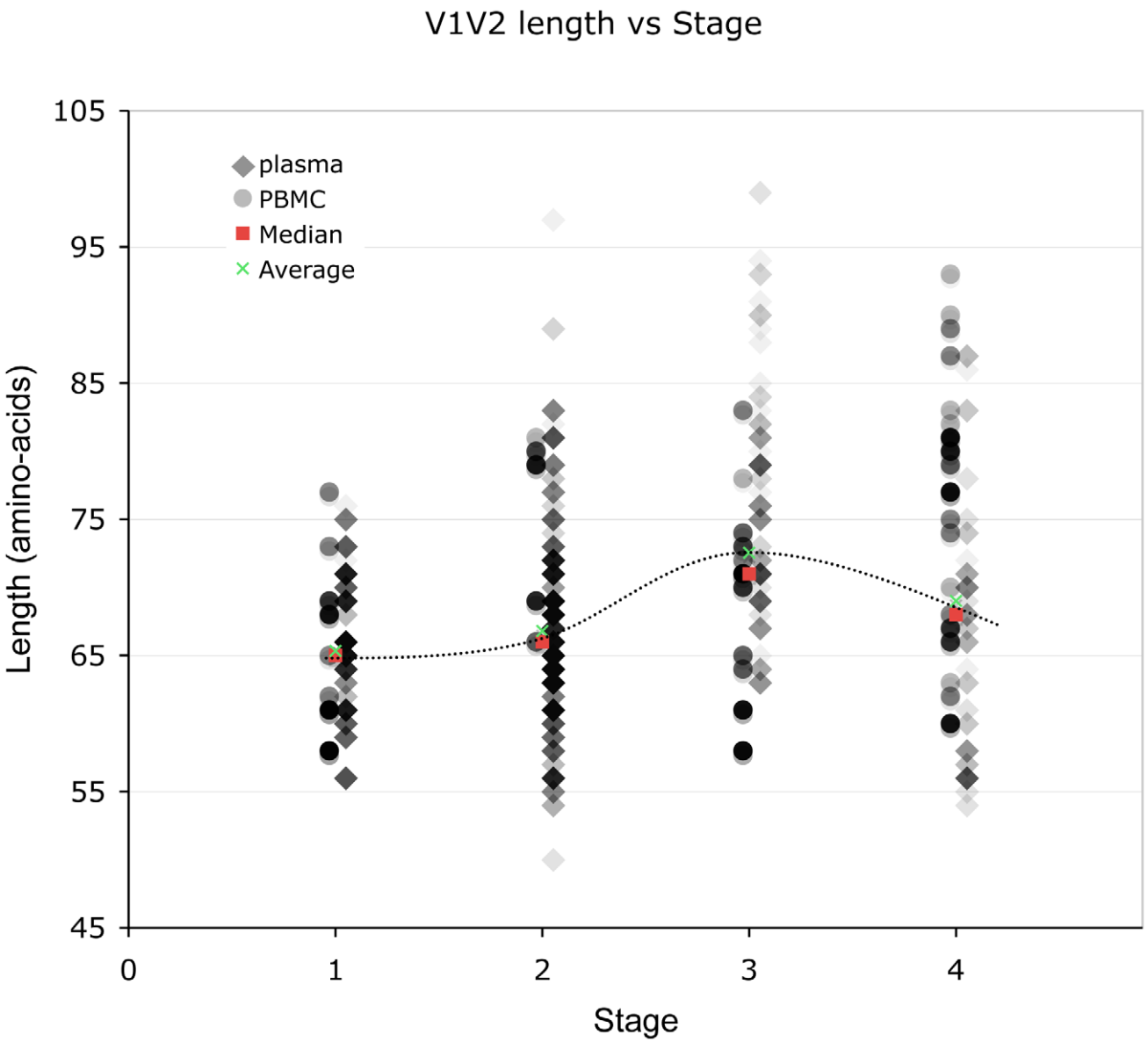


Figure 3. Correlation between stage of illness and V1V2 length. Lengths are indicated in amino acids. Sequences from plasma are represented by diamonds and sequences from PBMC are represented by circles. Overlapping data points appear as darker symbols. Quartiles and median values are indicated by horizontal line segments. Stage 1, 2, and 3 subjects were sampled within two months, between two months and three years, and at times >3 years post infection, respectively. Stage 4 subjects were comprised of all individuals meeting 1993 CDC criteria for AIDS when sampling occurred, regardless of time since infection.
doi:10.1371/journal.ppat.1001228.g003

Table 3. Multivariable regression analysis of V1V2 length vs. clinical variables.

	Time since infection (Model 1)			Infection Stage (Model 2)		
Factor	**All seqs**	**Plasma**	**PBMC**	**All seqs**	**Plasma**	**PBMC**
Time since infection	0.59 (0.85)	0.77 (<0.001)	0.56 (0.86)			
Stage 1				Ref	Ref	Ref
Stage 2				0.09 (0.9)	0.08 (0.9)	0.00 (0.9)
Stage 3				6.25 (<0.001)	6.65 (<0.001)	3.66 (0.10)
Stage 4				3.54 (0.02)	2.95 (0.06)	8.01 (0.01)
Factor	**CD4 count (Model 3)**			**Viral Load (Model 4)**		
CD4 level	**All seqs**	**Plasma**	**PBMC**	**All seqs**	**Plasma**	**PBMC**
>500	Ref	Ref	Ref			
200-500	0.87 (0.5)	0.45 (0.7)	8.13 (0.03)			
<200	−0.19 (0.9)	−0.70 (0.7)	6.77 (0.09)			
VL log10 (continuous)				−0.61 (0.14)	−0.58 (0.2)	−0.84 (0.15)

Beta coefficients for V1V2 Length vs. Time since Infection (Model 1), Stage of Infection (Model 2), CD4 counts (Model 3) or HIV Viral Load (Model 4). β values and p-values (in parentheses) are shown. Results are stratified by sample type (Plasma vs. PBMC), adjusting for year of sample collection. Time since infection was missing for 5 sequences, stage of infection for 242 sequences, CD4 count for 113 sequences, and viral load for 290 sequences with measured V1V2 length. Ref = Reference group. Analyses were performed for all sequences collectively as well as for sequences derived from plasma and PBMC considered separately.

doi:10.1371/journal.ppat.1001228.t003

resulting in 100 resampled datasets. These analyses confirmed that the observed relationship between V1V2 length and time since infection, and year of sampling were not significantly biased due to the inclusion of individuals with multiple sequences (See Text S1, section S2 and Figure S5).

Relationship between V4 and V5 loop lengths and clinical variables. (Also see Text S1, section S4 and Figure S7.) Despite their high degree of length variability, V4 and V5 loop lengths did not appear to vary significantly by time since infection in univariate regression analyses. In separate analyses adjusting for sample year and type, V4 length appeared somewhat increased in those with stage 2 ($\beta = 1.22$, $p = 0.02$), stage 3 ($\beta = 0.98$, $p = 0.09$) or stage 4 ($\beta = 1.11$, $p = 0.10$) compared to stage 1 infection. In contrast, V5 length decreased with increasing time after infection ($\beta = -0.07$, $p<0.001$), was decreased in stage 4 ($\beta = -0.69$, $p = 0.01$ compared to stage 1), and was decreased in those with CD4 counts below 200 cells/ml ($\beta = -0.66$, $p = 0.002$) compared to those with CD4 counts above 500 cells/ml.

Relationship between subregion glycosylation and clinical variables. (And see Text S1 section S1, and Figures S3, S4 and S8.) In separate univariate GEE analyses, the number of PNLGS in V1V2 increased with calendar year of sampling ($\beta = 0.06$ increase per year; $p = 0.02$), but was not significantly associated with sample type ($\beta = 0.32$ more potential sites in PBMC compared to plasma; $p = 0.17$). Glycosylation in V1V2 was increased in those with stage 3 ($\beta = 0.96$, $p = 0.002$), but not stage 2 or 4 compared to stage 1 infection, and was decreased in those with CD4 counts <200 cells/ml ($\beta = -0.63$, $p = 0.04$ compared to CD4 counts >500). Similar findings were obtained in an analysis restricted to sequences derived from plasma; the number of PNLGS in V1V2 increased with calendar year of sampling ($\beta = 0.05$ increase per year; $p = 0.001$), was increased in those with stage 3 infection ($\beta = 1.14$, $p = 0.001$) and was decreased in those with CD4 counts <200 cells/ml ($\beta = -0.84$, $p = 0.04$ compared to CD4 counts >500). However, in PBMC, the number of sequences was limited, and no associations between the number of potential glycosylation sites and clinical features achieved statistical significance. Glycosylation in V4 decreased ($p<0.001$), while in V5 glycosylation increased with calendar year ($\beta = 0.01$ per year, $p = 0.02$), although the magnitude of these effects was small ($\beta = -0.03$ per year, and 0.01 per year, respectively).

Coreceptor usage, clinical factors and V1V2 loop length. (Also see Text S1 section S5 and Figures S10 and S11.) We next used four published genotypic methods to infer coreceptor usage based on V3 loop amino acid sequence [62–65]. In our dataset, 4476 V3 loop sequences were available for scoring, and were derived from 129 individuals. 121 V3 loops could not be scored by the PGRC method because the aligned sequences exceeded the length limit specified by the input format (40 characters). There was agreement in coreceptor usage assignment by all of the methods in 3644 of 4476 sequences (81.4%) and disagreement between one or more methods in the remaining 832 sequences. 1046 of 4476 sequences were scored as CXR4-using or syncytium-inducing by one or more methods, and the remaining 3430 were uniformly scored as CCR5 or non-syncytium by all methods. 60 of 129 individuals had at least one X4-scoring V3 loop as determined by one or more of the prediction methods, while the remaining 69 had only CCR5-scoring sequences.

We then considered inferred coreceptor usage as a function of time since infection, clinical stage, CD4 counts, HIV viral load, and V1V2 length, both overall and separately in plasma- and PBMC–derived viruses. Because the PSSM method provides a continuous numerical measure corresponding to the sequence position on a continuum of the evolutionary changes leading to X4 usage (the PSSM score), we examined PSSM score in relation to these variables. Overall, in separate GEE regression analyses, PSSM score was not related to time since infection ($p = 0.9$) or HIV viral load ($p = 0.5$). However, PSSM score was significantly increased (indicating greater CXCR4 usage) in those with stage 4 ($\beta = 6.34$, $p = 0.0002$) but not stage 2 or stage 3 infection ($p = 0.8$ each). Similarly, PSSM score was significantly increased in those with intermediate (200–500 cells/ml) and low (<200 cells/ml) CD4 counts ($\beta = 1.52$, $p = 0.02$ and $\beta = 6.62$; $p<0.0001$, respectively) compared to those with CD4 counts above 500 cells/ml. PSSM score was weakly associated with increased V1V2 length

($\beta = 0.06$; $p = 0.09$ per one amino acid increase in V1V2 length). The analyses restricted to plasma samples yielded similar results, with PSSM score strongly associated with stage 4 infection ($\beta = 8.54$, $p<0.0001$), intermediate (200–500 cells/ml) and low (<200 cells/ml) CD4 counts ($\beta = 1.67$, $p = 0.03$ and $\beta = 8.83$; $p<0.0001$, respectively) compared to those with CD4 counts above 500 cells/ml, and PSSM score weakly associated with increased V1V2 length ($\beta = 0.07$; $p = 0.12$ per one amino acid increase in V1V2 length). In sequences derived from PBMC samples, PSSM score was not associated with stage of infection, CD4 counts, HIV viral load, or V1V2 length. However, PSSM score was inversely associated with time since infection ($\beta = -0.15$; $p = 0.01$ per year).

Longitudinal analyses

In the longitudinal dataset, significant V1V2 length increases between first and second timepoints were noted in 10 of 22 subjects, a significant V1V2 length decrease over time occurred in one subject, and no significant V1V2 length changes over time were seen in the remaining 11 subjects. These findings appeared to vary by stage of infection (t-test $p = 0.03$). In the 15 patients from the L1 group (individuals not meeting AIDS criteria at any time prior to final sampling), the mean increase of V1V2 length per subjects was 1.69 amino acids per year, and 9 subjects experienced significant V1V2 length increases over time (Figures 4 and 5). In contrast, of the seven subjects in the L2 group (individuals progressing to AIDS between first and final sample), the mean V1V2 length decreased by an average of 0.10 amino acids per year, with only one having a significant trend of increasing length, while one individual showed a significant decrease in length (Figure 6). The distribution of V1V2 length change (increase or decrease) by group was therefore asymmetric (Fisher's exact test, $p = 0.02$), reflecting a trend of increasing length in asymptomatic individuals (group L1) and stable or decreasing length in individuals with AIDS (group L2) (Table 4). Three subjects in group L1 had extensive longitudinal sampling (Figure 5); in 1362 and Q23 [51], there was a period of V1V2 length stability of approximately 2 years, followed by increase through 4.5 years. V1V2 length increase over time was also seen in CC1. In the case of CC1, a pseudotyped virus was created using the gp120 coding region from the initial timepoint from this individual in a HIV-1 NL4-3 background, and cultured *in vitro* [54]. In contrast to the patterns observed *in vivo*, V1V2 length and number of glycosylation sites both declined rapidly over 20 generations *in vitro* ($p<0.001$).

Discussion

We have systematically examined gp120 subregion length variation, and the relationship between length polymorphism, N-linked glycosylation sites, and clinical markers of disease progression. Although V1V2, V4 and V5 all displayed remarkable length heterogeneity, and V1V2, C3 and V4 were also quite variable with respect to glycosylation, the most significant associations between virological and clinical variables localized to the V1V2 region. We found that V1V2 length and glycosylation increased significantly over time during chronic infection, and

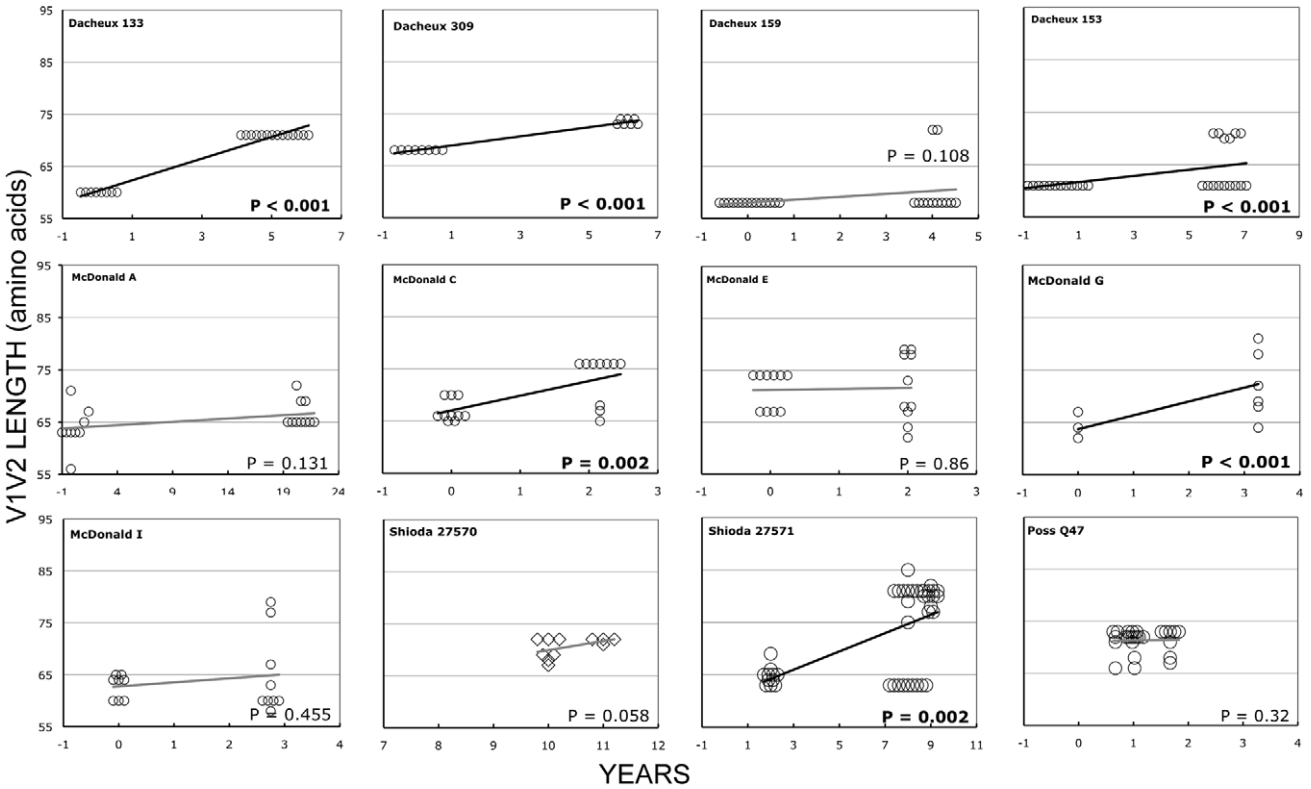
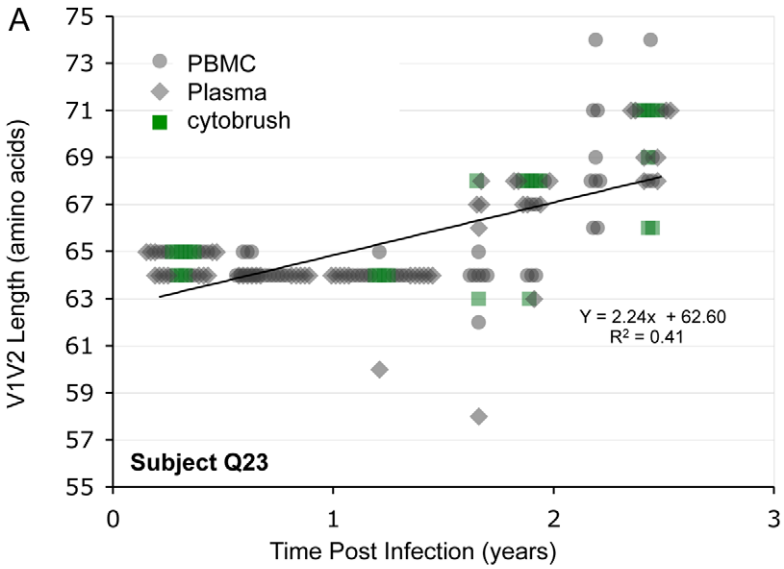
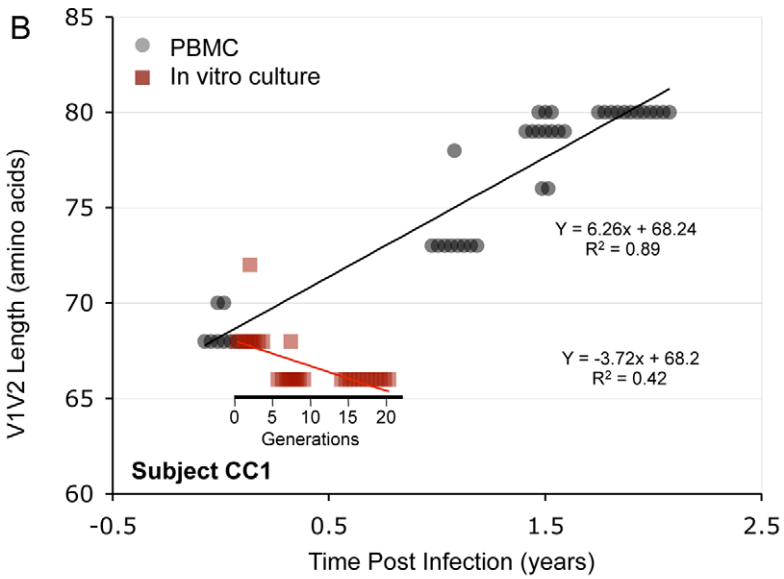


Figure 4. V1V2 loop lengths over time in group L1. Sequences from plasma are represented by diamonds and sequences from PBMC are represented by circles. Significant slopes are indicated in bold. X-axis denotes years elapsed between sampling time points, but do not necessarily indicate the total duration of infection. The first author of the report in which data were originally presented is indicated in the upper left-hand corner of each graph. Group L1 subjects did not meet criteria for AIDS at any time prior to the final sample. Subjects reported by McDonald et al had received AZT monotherapy at one or more times prior to sampling.
doi:10.1371/journal.ppat.1001228.g004


A
PBMC
Plasma
cytobrush
V1V2 Length (amino acids)
Y = 2.24x + 62.60
R² = 0.41
Subject Q23
Time Post Infection (years)


B
PBMC
In vitro culture
V1V2 Length (amino acids)
Y = 6.26x + 68.24
R² = 0.89
Y = -3.72x + 68.2
R² = 0.42
Generations
Subject CC1
Time Post Infection (years)

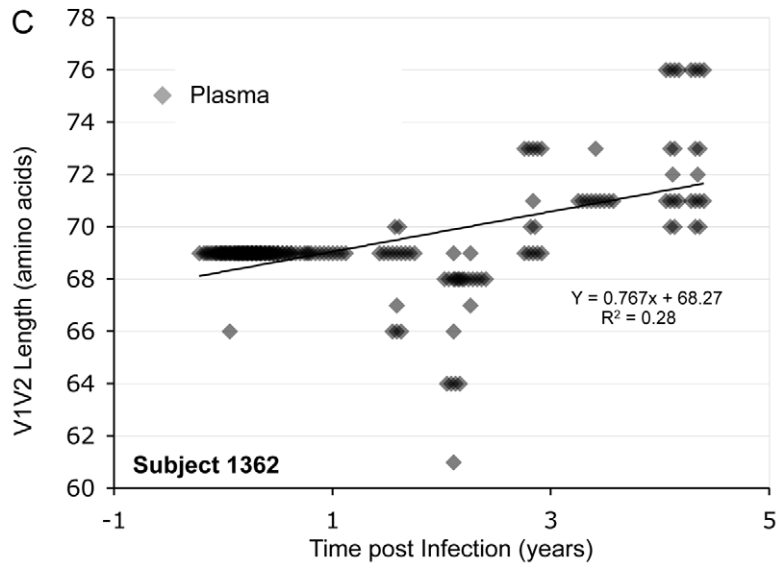

C
Plasma
V1V2 Length (amino acids)
Y = 0.767x + 68.27
R² = 0.28
Subject 1362
Time post Infection (years)

Figure 5. V1V2 length *vs.* time in subjects Q23, CC1 and 1362. Panel A: Subject Q23, infected with HIV subtype A. Sequences were derived from PBMC (black circles), plasma (black diamonds) and DNA from cervical lymphocytes (green squares) as described by Poss et al [80]. Panel B: Subject CC1, infected with subtype A. Sequences were obtained from plasma (black diamonds) and tissue culture (red squares). Length change of *in vitro* sequences occurs over ~ 40 days, and are represented along an expanded X-axis for clarity.
doi:10.1371/journal.ppat.1001228.g005

then declined in late-stage illness. In regression analyses, time since infection was the most influential factor in determining V1V2 length. In addition, there was a modest but significant increase in V1V2 length over the period from 1984–2004. V5 loop length was highly variable, but tended to decrease slightly in length over the course of infection.

In SIV infection, the number of PNLGS in gp120 increases over time *in vivo* following inoculation of a cell-passaged strain [71]. In one earlier study in humans, Bunnik et al noted expansion in gp120 length followed by contraction over time in 4 of 5 individuals receiving antiretroviral therapy, and similar changes in glycosylation in 3 subjects [72]. Others have noted a relationship between early infection and reduced V1V2 length and glycosylation in subtypes C and A [19,45]. In contrast, a comparison of early and chronic HIV-1 subtype B sequences from the HIV sequence database failed to reveal any significant difference in V1V2 length [19], suggesting that these effects may be subtype-specific. Data on length/glycosylation changes during transmission have been conflicting. Derdeyn et al [45] demonstrated reduced length and glycosylation in V1–V4 following heterosexual transmission in HIV-1 subtype C. However, Frost et al failed to note similar findings in a study of eight subtype B homosexual transmission pairs [47], and in our examination of these and 10 additional subtype B infected homosexual transmission pairs, we found no consistent pattern of change in V1–V2 or V1–V4 length or glycosylation upon transmission [46].

Interpretation of the data presented here may be affected by several methodological factors. There is probably some variation in the accuracy of the reported time of infection for sequences obtained from previous reports. In some cases, sequences obtained from prior publications may have been obtained under conditions permitting template resampling [73], and a systematic error due to evolving laboratory methods could result in bias. Also, in our analyses, we have not formally corrected for multiple comparisons. Physiological factors are also likely to introduce some noise, particularly in cross-sectional analyses of parameters with respect to time since infection. The individuals included here represent a broad spectrum of clinical scenarios, diverse host immune response profiles and varying disease progression rates. Plasma sequences may receive contributions from both recently infected target cells and older reservoirs, and therefore imperfectly reflect selective pressures prevailing at the time of infection. Finally, length and glycosylation phenotypes are likely to be affected by chance events and unknown factors not considered in our analyses. Therefore, the effects we describe are influential rather than deterministic, and reflect important selective forces that can be discerned against a background of high inter-individual variation.

Despite these limitations, the analyses presented here and the work of others [40,45–47,72] provide the outlines of an overall pattern characterized by transmission of randomly selected V1V2 loop lengths from viruses present in the donor pool, a brief decline in loop size during the initial months immediately following infection, gradual selection for bulkier V1V2 loops during chronic infection, and finally, reversion to more compact loops during late stage illness. Structural studies [22,23], neutralization studies [20,34–38,42], and *in vitro* data on viruses lacking V1 and V2 [43,44] suggest that one major function of the V1V2 region may be to permit evasion from humoral immune responses in the host. Thus, the trends outlined above support the hypothesis that HIV populations may evolve to escape humoral selective pressure by increasing V1V2 loop size. According to this view, the newly infected, immunologically naïve host might be expected to harbor relatively short V1V2 loops that eventually lengthen in response to

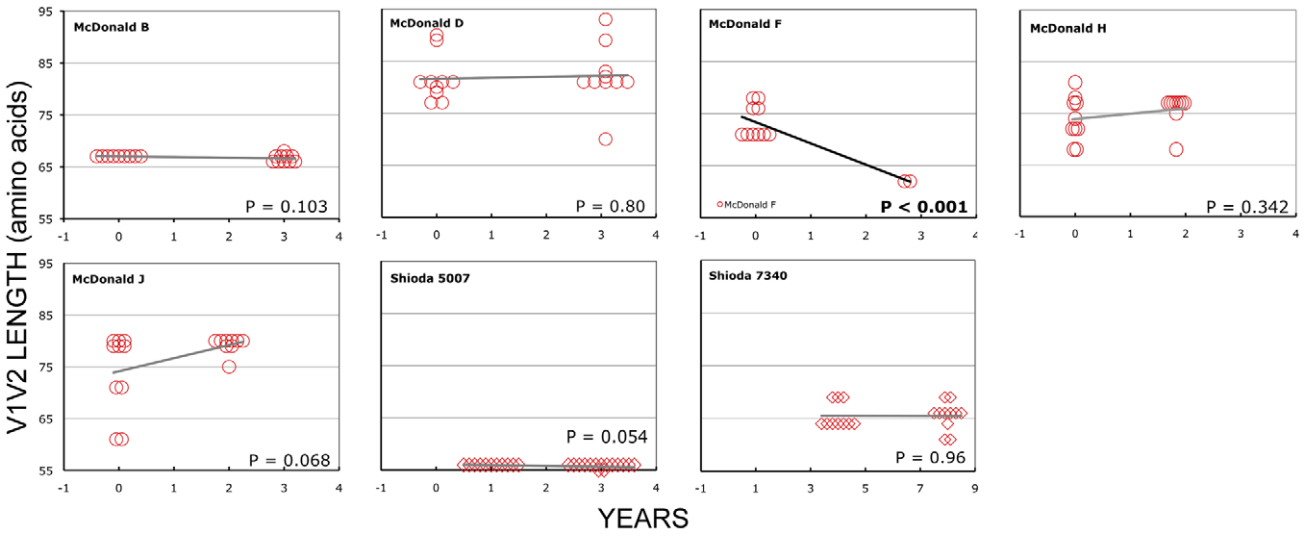


Figure 6. V1V2 loop lengths over time in group L2. Sequences from plasma are represented by diamonds and sequences from PBMC are represented by circles. Significant slopes are indicated in bold. X-axis denotes years elapsed between sampling time points, but do not necessarily indicate the total duration of infection. The first author of the report in which data were originally presented are indicated in the upper left-hand corner of each graph. Group L2 subjects were reported to have an AIDS-defining illness or peripheral CD4 count $<200/mm^3$ between the first and second samples.
doi:10.1371/journal.ppat.1001228.g006

Table 4. Summary of longitudinal data.

ID	Slope	p-value	Group
133	2.1774	**<.0001**	L1
1362	0.7478	**<.0001**	L1
153	0.6097	**0.0016**	L1
159	0.5801	0.0943	L1
27570	1.8929	0.0566	L1
27571	1.8123	**<.0001**	L1
309	0.8926	**<.0001**	L1
A	1.25	0.0575	L1
C	2.8704	**<.0001**	L1
CC1	6.29	**<.0001**	L1
E	0.2	0.8591	L1
G	2.6462	**<.0001**	L1
I	0.7778	0.4324	L1
Q23	2.4755	**<.0001**	L1
Q47	0.1087	0.808	L1
5007	−0.1765	0.0856	L2
7340	−0.0114	0.9655	L2
B	−0.1333	0.0704	L2
D	0.1948	0.788	L2
F	−4.1455	**<.0001**	L2
H	1.0261	0.2487	L2
J	2.5611	**0.0401**	L2

Initial parameter estimates and p-values were used when only two time points were available. When multiple timepoints were available, final GEE estimates and p-values were used. P-values less than 0.05 are shown in bold.
doi:10.1371/journal.ppat.1001228.t004

an effective humoral response at some fitness cost (Figure S9). Experimental evidence indicating that relaxation of antibody-mediated selective pressure during early infection is associated with shorter loops is provided by Derdeyn, who demonstrated significantly greater neutralization sensitivity among five recipients during early infection, than in the corresponding donors [45]. The decline in V1V2 size observed in advanced disease probably reflects waning effectiveness of humoral immunity in hosts with late-stage illness and profound immune dysregulation (Figure 7). This decline is also congruent with previous findings of an inverse relationship between the rate of HIV genetic evolution and the rate of CD4 T cell decline in some individuals [74]. The dramatic reduction in V1V2 length associated with transfer to the *in vitro* environment [54] represents the extreme case of absent host immunity, where viruses without an unnecessarily bulky V1V2 loop achieve maximum replicative fitness. As would be expected, the patterns we observe are most pronounced in plasma sequences, which most directly reflect the selective forces present at the time of sampling. In contrast, a significant increase in V1V2 length over time was not seen in the PBMC compartment. These observations are consistent with the presence of archived genotypes from earlier times during the course of infection within the PBMC compartment. We also note that genotypes present in plasma may emanate from other cellular compartments in addition to PBMC, and may therefore reflect somewhat different evolutionary pressures. However, a considerably greater number of V1V2 sequences were derived from plasma, and sample size may also account for some of the differences observed between these compartments.

Our model may help to explain a failure to find any significant difference in V1V2 length in a comparison of early and chronic HIV-1 subtype B sequences (including sequences from late-stage individuals) [19]. When we reanalyzed the data presented by Chohan [19] after separating subjects with stable chronic illness from subjects with AIDS (Figure S13), we observed a pattern of

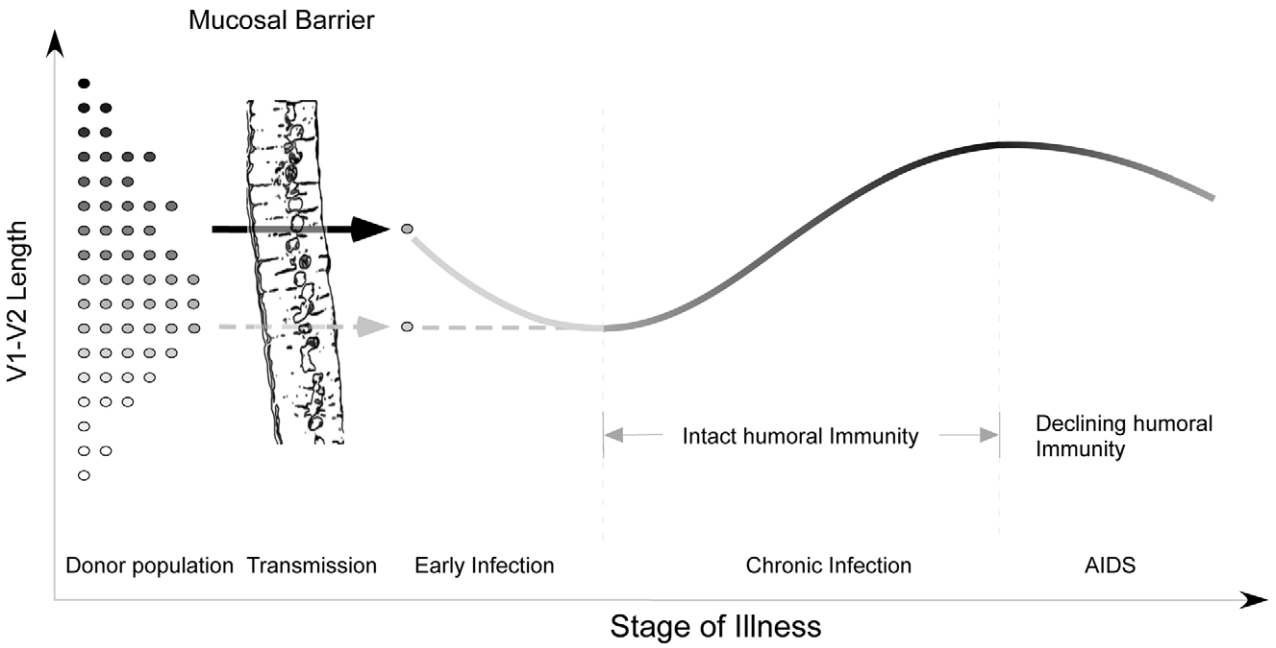


Figure 7. Proposed evolution of V1V2 loop size change during transmission and HIV infection. At the time of sexual transmission, a significant genetic bottleneck occurs in which one or a small number of donor variants is transmitted to the recipient, without clear selection for loop size (represented on the y-axis). During early infection, prior to an effective host response, viral variants with a compact V1V2 loop have a competitive advantage, and V1V2 loop size remains stable or regresses. During chronic asymptomatic infection, mean V1V2 length increases in response to (humoral) immune selective pressure. As immune function wanes, V1V2 loop length gradually declines.
doi:10.1371/journal.ppat.1001228.g007

lengthening over time, followed by decline in late-stage illness, as reported here (See Text S1, section S7). Similarly, we may explain discordant results obtained on V1V2 length variation during transmission of HIV-1 subtypes C and B. While a trend towards shorter loops in recipients was seen in subtype C [45] but not B [46,47], it is likely for methodological reasons that the subjects studied by Derdeyn were sampled at somewhat later times than those of Frost and Liu. Thus the sequences in the latter two studies would be expected to be a random sampling from the donor pool, while those of Derdeyn might reflect the expected shortening prior to the onset of an effective antibody response. Indeed, when we examine a much larger set of subtype A and C transmission pairs from East Africa with more precisely known sampling times obtained soon after transmission, it is difficult to appreciate any consistent pattern of V1V2 length change (See Text S1, section S8 and Figure S14). Thus there may be no need to infer separate mechanisms for different HIV-1 subtypes and modes of transmission.

In addition, we may also explain a trend of increasing V1V2 length by calendar year. If shorter and less glycosylated V1V2 were always selected during transmission, transmission from donors in early infection would maintain a constant V1V2 length within the epidemic, whereas if all new cases were acquired from chronically infected hosts, this increase of V1V2 length by calendar year could be dramatic. However, most studies suggest that about half of transmission events involve subjects in early infection [46,75,76], consistent with the moderate trend we observed. Alternatively, the temporal trends we have observed could represent a gradual adaptation by HIV-1 to host the host environment at the population level, a hypothesis that has been proposed by several investigators with respect to mutational escape from HLA-restricted CTL epitopes [77–79].

Finally, our results imply that the polymorphisms seen in V1V2 reflect the ability of the host to mount a meaningful immunological response, rather than virologic features that dictate the course of illness. That is, we argue that V1V2 length change is a consequence of environmental selective pressure rather than a causative factor in disease progression.

Supporting Information

Text S1 Supporting analyses text - PDF document containing text containing supplementary analyses and citations.
Found at: doi:10.1371/journal.ppat.1001228.s001 (0.14 MB PDF)

Figure S1 V1V2 length vs. virologic and clinical parameters I. **Panel A:** V1V2 length vs. $\log_{10}$ plasma viral load (no significant relationship). **Panel B:** V1V2 length vs. peripheral CD4 T-cell count (no significant relationship). **Panel C:** V1V2 length by coreceptor usage. Box-plots depict minimum, 1st quartile, median (red line), 3rd quartile and maximum values in each group, with superimposed individual length measurements. In this series, V1V2 sequences associated with V3 loops predicted to be X4-tropic by PSSM are slightly longer compared with sequences associated with R5-tropic V3 loops (median 71 vs. 66 amino acids, $p = 3.49\times10^{-5}$, MW test). However, a plot of V1V2 length vs. PSSM score (**Panel D**) does not reveal a clear linear correlation between V1V2 length and PSSM score.
Found at: doi:10.1371/journal.ppat.1001228.s002 (1.15 MB TIF)

Figure S2 V1V2 length vs. virologic and clinical parameters II. **Panel A:** V1V2 length vs. time since infection. As described earlier, a significant positive correlation V1V2 length and time since infection is evident ($R = 0.149$) **Panel B:** V1V2 length by stage of infection. Box-plots depict minimum, 1st quartile, median (red line), 3rd quartile and maximum values in each stage group, with superimposed individual length measurements. Highly significant differences in V1V2 length are seen between stage 3 and stages 1,2 and 4 ($p<2.2\times10^{-16}$, M-W rank sum test), reflecting V1V2 lengthening in chronic illness, followed by contraction in late disease. **Panel C:** V1V2 length by site (PBMC vs. plasma). In this univariate comparison, there is no significant length difference between V1V2 loops obtained from PBMC (median 68 amino acids) and plasma (median 66 amino acids, $p = 0.93$). **Panel D:** V1V2 length vs. year of sampling. As described, there is a significant positive correlation between V1V2 length and year of sampling.
Found at: doi:10.1371/journal.ppat.1001228.s003 (1.14 MB TIF)

Figure S3 V1V2 glycosylation vs. virological and clinical parameters I. **Panel A:** Number of V1V2 glycosylation sites vs. $\log_{10}$ plasma viral load (no significant relationship). **Panel B:** Number of V1V2 glycosylation sites vs. peripheral CD4 T-cell count (no clear correlation observed). **Panel C:** V1V2 glycosylation sites by inferred coreceptor usage (R5 or X4). Box-plots report minimum, 1st quartile, median (red line), 3rd quartile and maximum values in each stage group, with superimposed individual measurements. No clear differences in glycosylation are noted between V1V2 loops associated with R5-tropic and X4-tropic V3 loops (median 6 and 6 PNLGS, respectively). **Panel D:** Number of V1V2 glycosylation sites vs. PSSM score (no significant relationship).
Found at: doi:10.1371/journal.ppat.1001228.s004 (1.07 MB TIF)

Figure S4 V1V2 glycosylation vs. virological and clinical parameters II. **Panel A:** Number of V1V2 glycosylation sites vs. time since infection. As with V1V2 length, in a univariate analysis there is a modest but significant linear correlation between time since infection and the extent of V1V2 glycosylation ($\beta = 0.12$ amino acids/year, $R^2 = 0.09$). **Panel B:** Number of V1V2 glycosylation sites by clinical stage. Similar to what was observed for V1V2 length, glycosylation in chronic illness (stage 3) was significantly greater than in early and late disease ($p<1\times10^{-8}$), reflecting increasing glycosylation during chronic infection, followed by a decline in the extent of glycosylation during AIDS. **Panel C:** V1V2 glycosylation sites by site (PBMC or plasma). Box-plots report minimum, 1st quartile, median (red line), 3rd quartile and maximum values in each stage group, with superimposed individual measurements. No clear differences in glycosylation are noted between V1V2 loops obtained from PBMC vs. plasma (median PNLGs 5 and 6, respectively, $p = 0.59$). **Panel D:** Number of V1V2 glycosylation sites vs. year of sampling. There is a negligible positive correlation between V1V2 PNLG and year of sampling ($\beta = 0.05$, $R^2 = 0.06$)
Found at: doi:10.1371/journal.ppat.1001228.s005 (1.13 MB TIF)

Figure S5 Resampling analysis: R^2 values for multiple linear regression of V1V2 length on the independent variables time since infection, year of sampling, and sample type for the entire dataset (red squares □) and for 100 parallel randomly resampled datasets derived from the original dataset (green diamonds ◇). Correlation coefficients obtained in the resampled datasets were consistent with the correlation coefficient obtained using all data.
Found at: doi:10.1371/journal.ppat.1001228.s006 (1.33 MB TIF)

Figure S6 V1V2 sequence length vs. time since infection - sliding window analysis: Length measurements (red + sign) and R^2 values (blue triangles Δ) for univariate linear regression analyses of datasets excluding 0.4-year periods since the time of infection. 0.4-

year data exclusion periods are centered around the x value of each Δ datapoint. The correlation strength of the linear model is greatest for datasets excluding the earliest two 0.4-year periods (first two datapoints), indicating that linear regression of V1V2 length on time since infection most accurately explains data obtained at times after approximately 0.8 years.
Found at: doi:10.1371/journal.ppat.1001228.s007 (1.07 MB TIF)

Figure S7 Subregion length (V1–V5) vs. time since infection for the V1V5 region (purple circles), the V1V4 region (green diamonds) and V1V2 (blue squares), V4 (red triangles) and V5 (orange diamonds) considered separately. A significant trend towards increasing length seen in V1V2, V1V4 and V1V5 can be ascribed primarily to changes in V1V2.
Found at: doi:10.1371/journal.ppat.1001228.s008 (2.13 MB TIF)

Figure S8 Subregion glycosylation (V1–V5) vs. time since infection for the V1V5 region (purple circles), the V1V4 region (green diamonds) and for V1V2 (blue squares) and V4 (red triangles) considered separately. A modest trend towards increasing glycosylaton seen in V1V2, V1V4 and V1V5 can be ascribed primarily to changes in V1V2
Found at: doi:10.1371/journal.ppat.1001228.s009 (2.38 MB TIF)

Figure S9 Agreement between 4 bioinformatic coreceptors used to assign probable coreceptor usage. There was complete agreement between all methods for ~80% of sequences examined, while in the remaining 20%, there was some disagreement in assignment between one or more scoring methods. Most sequences were predicted to be CCR5-tropic by all methods (white bar), while a modest number of sequences was predicted to be CXCR4-tropic by all methods. The remaining sequences were scored differently by various methods, as represented (colored bars).
Found at: doi:10.1371/journal.ppat.1001228.s010 (2.84 MB TIF)

Figure S10 V1V2 sequence length *vs.* time since infection and PSSM score. Rising PSSM scores (color scale), depicted as warmer colors, indicate a greater likelihood of CXCR4 coreceptor usage; in this dataset, predicted X4 coreceptor usage occurs at a PSSM score of approximately -2. In these data, there is a pronounced preponderance of CCR5-using viruses, with a trend towards increasing prevalence of X4-tropic viruses during chronic infection. However, X4 and R5 viruses are distributed throughout all infection times, and cannot be easily distinguished on the basis of V1V2 length.
Found at: doi:10.1371/journal.ppat.1001228.s011 (1.15 MB TIF)

Figure S11 V1V2 potential N-linked glycosylation sites *vs.* V1V2 length and PSSM score (color scale). There is a very marked dependence of glycosylation on length ($\beta = 0.13$ PNGL/amino acid, $R^2 = 0.52$). X4-usage appears to be more commonly associated with V1V2 sequences bearing 4-7 PNLG sites, than with sequences with more than 7 sites (and see figure S1 panel D).
Found at: doi:10.1371/journal.ppat.1001228.s012 (1.10 MB TIF)

Figure S12 V1V2 and Stage of Illness. V1V2 length vs. Time since Infection for stage 1 (orange "+"), stage 2 (gray triangles), stage 3 (blue squares), and stage 4 (red diamonds). There is a slight decline in V1V2 length from stage 1 to stage 2, reflecting regression from transmitted viruses of essentially random lengths to shorter loop lengths during early infection prior to the onset of a meaningful immune response. This is followed by a strong trend towards lengthening during chronic infection (stage 3) and a weakening of this trend in late-stage illness (stage 4).
Found at: doi:10.1371/journal.ppat.1001228.s013 (1.52 MB TIF)

Figure S13 Chohan Data revisited: V1V2 sequence length for subjects in early infection (first bar), chronic infection and AIDS considered together (second bar), chronic stable infection only (third bar), and individuals with AIDS-defining clinical conditions (fourth bar). Length differences between "early", "chronic" and "AIDS" are statistically significant ($p \leq 0.02$). Thus, separation of sequences obtained during AIDS from sequences obtained during chronic stable infection reveals a trend of rising V1V2 length through chronic infection, followed by falling length in AIDS that is not otherwise apparent.
Found at: doi:10.1371/journal.ppat.1001228.s014 (1.04 MB TIF)

Figure S14 V1V2 length during transmission: Change in mean loop length between donor and recipient in 44 transmission pairs involving HIV-1 subtypes A, C and B, presented by Haaland, Derdeyn, Frost and Liu. **Panels A–C**: Difference in mean loop length between donors and recipients vs. time since infection for V1V2 (**panel A**), C2–V4 (**panel B**), and V1–V4 (**panel C**). **Panel D**: Difference in mean loop length between donors and recipients vs. the mean loop length (for the corresponding region) in the donor. Subtype A sequences (Haaland, represented by red +), Subtype B sequences (Frost, blue X) and Liu (blue squares) and subtype C sequences (Haaland, green squares, and Derdeyn, green X).
Found at: doi:10.1371/journal.ppat.1001228.s015 (2.25 MB TIF)

Table S1 Published sequence data. Accession Numbers for previously published sequences included in cross-sectional, longitudinal and transmission analyses.
Found at: doi:10.1371/journal.ppat.1001228.s016 (0.05 MB PDF)

Table S2 Multivariable regression analysis of V1V2 length vs. clinical variables, upper and lower 5% excluded. Beta coefficients for V1V2 Length vs. Time since Infection (Model 1), Stage of Infection (Model 2), CD4 counts (Model 3) or HIV Viral Load (Model 4). β values and p-values (in parentheses) are shown. Results are stratified by sample type (Plasma vs. PBMC), adjusting for year of sample collection. Time since infection was missing for 5 sequences, stage of infection for 242 sequences, CD4 count for 113 sequences, and viral load for 290 sequences with measured V1V2 length. Ref = Reference group. Analyses were performed for all sequences collectively as well as for sequences derived from plasma and PBMC considered separately. Sequences comprising the upper and lower 5% by length were excluded from these analyses.
Found at: doi:10.1371/journal.ppat.1001228.s017 (0.06 MB PDF)

Acknowledgments

We would like to thank Drs. Cynthia Derdeyn, Eric Hunter, Simon Frost, Serena Spudich, Richard Price, Patrizia Bagnarelli and Eric Delwart for their assistance and helpful discussions during the collection and analysis of these data.

Author Contributions

Conceived and designed the experiments: MEC RZ. Performed the experiments: MEC. Analyzed the data: MEC RZ SEH YL WD GSG TZ JIM. Contributed reagents/materials/analysis tools: TZ JIM. Wrote the paper: MEC RZ SEH YL WD GSG TZ JIM.

References

1. Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, et al. (1986) Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell 45: 637–648.

2. Willey RL, Rutledge RA, Dias S, Folks T, Theodore T, et al. (1986) Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. Proc Natl Acad Sci USA 83: 5038–5042.
3. Modrow S, Hahn BE, Shaw GM, Gallo RC, Wong-Staal F, et al. (1987) Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions. J Virol 61: 570–578.
4. Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, et al. (2009) HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. PLoS Pathog 5: e1000414.
5. Cocchi F, DeVico AL, Garzino-Demo A, Cara A, Gallo RC, et al. (1996) The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockade of infection [see comments]. Nat Med 2: 1244–1247.
6. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272: 872–877.
7. Speck RF, Wehrly K, Platt EJ, Atchison RE, Charo IF, et al. (1997) Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. J Virol 71: 7136–7139.
8. Goudsmit J, Debouck C, Meloen RH, Smit L, Bakker M, et al. (1988) Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. ProcNatlAcadSciUSA 85: 4478–4482.
9. Javaherian K, Langlois AJ, McDanal C, Ross KL, Eckler LI, et al. (1989) Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. ProcNatlAcadSciUSA 86: 6768–6772.
10. Luo L, Li Y, Chang JS, Cho SY, Kim TY, et al. (1998) Induction of V3-specific cytotoxic T lymphocyte responses by HIV gag particles carrying multiple immunodominant V3 epitopes of gp120. Virology 240: 316–325.
11. Watanabe N, McAdam SN, Boyson JE, Piekarczyk MS, Yasutomi Y, et al. (1994) A simian immunodeficiency virus envelope V3 cytotoxic T-lymphocyte epitope in rhesus monkeys and its restricting major histocompatibility complex class I molecule Mamu-A*02. J Virol 68: 6690–6696.
12. Hartley O, Klasse PJ, Sattentau QJ, Moore JP (2005) V3: HIV's switch-hitter. AIDS Res Hum Retroviruses 21: 171–189.
13. Hill MD, Lorenzo E, Kumar A (2004) Changes in the human immunodeficiency virus V3 region that correspond with disease progression: a meta-analysis. Virus Res 106: 27–33.
14. Ida S, Gatanaga H, Shioda T, Nagai Y, Kobayashi N, et al. (1997) HIV type 1 V3 variation dynamics *in vivo*: Long-term persistence of non-syncytium-inducing genotypes and transient presence of syncytium-inducing genotypes during the course of progressive AIDS. AIDS Res and Human Retrovir 13: 1597–1609.
15. Palmer C, Balfe P, Fox D, May JC, Frederiksson R, et al. (1996) Functional characterization of the V1V2 region of human immunodeficiency virus type 1. Virology 220: 436–449.
16. Masciotra S, Owen SM, Rudolph D, Yang C, Wang B, et al. (2002) Temporal relationship between V1V2 variation, macrophage replication, and coreceptor adaptation during HIV-1 disease progression. Aids 16: 1887–1898.
17. Kitrinos KM, Hoffman NG, Nelson JA, Swanstrom R (2003) Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection. J Virol 77: 6811–6822.
18. Shioda T, Oka S, Xin X, Liu H, Harukuni R, et al. (1997) In vivo sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression. J Virol 71: 4871–4881.
19. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, et al. (2005) Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol 79: 6528–6531.
20. Sagar M, Wu X, Lee S, Overbaugh J (2006) Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. J Virol 80: 9586–9598.
21. Chackerian B, Rudensey LM, Overbaugh J (1997) Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. J Virol 71: 7719–7727.
22. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393: 648–659.
23. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, et al. (2005) Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. Structure 13: 197–211.
24. Cartier L, Hartley O, Dubois-Dauphin M, Krause KH (2005) Chemokine receptors in the central nervous system: role in brain inflammation and neurodegenerative diseases. Brain Res Brain Res Rev 48: 16–42.
25. Andeweg A, Leeflang P, Osterhaus A, Bosch M (1993) Both the V2 and V3 regions of the human immunodeficiency virus type 1 surface glycoprotein functionally interact with other envelope regions in syncytium formation. J Virol 67: 3232–3239.
26. Groenink M, Fouchier RAM, Broersen S, Baker CH, Koot M, et al. (1993) Relation of phenotype evolution of HIV-1 to envelope V2 configuration. Science 260: 1513–1516.
27. Koito A, Harrowe G, Levy JA, Cheng-Mayer C (1994) Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization. J Virol 68: 2253–2259.
28. O'Brien WA, Koyanagi Y, Namazie A, Zhao JQ, Diagne A, et al. (1990) HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature 348: 69–73.
29. Sullivan N, Thali M, Furman C, Ho DD, Sodroski J (1993) Effect of amino acid changes in the V1/V2 region of the human immunodeficiency virus type 1 gp120 glycoprotein on subunit association, syncytium formation, and recognition by a neutralizing antibody. J Virol 67: 3674–3679.
30. Westervelt P, Trowbridge DB, Epstein LG, Blumberg BM, Li Y, et al. (1992) Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J Virol 66: 2577–2582.
31. Toohey K, Wehrly K, Nishio J, Perryman S, Chesebro B (1995) Human immunodeficiency virus envelope V1 and V2 regions influence replication efficiency in macrophages by affecting virus spread. Virology 213: 70–79.
32. Wang N, Zhu T, Ho DD (1995) Sequence diversity of V1 and V2 domains of gp120 from human immunodeficiency virus type 1: lack of correlation with viral phenotype. J Virol 69: 2708–2715.
33. Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, et al. (2006) Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol 80: 750–758.
34. Benichou S, Legrand R, Nakagawa N, Faure T, Traincard F, et al. (1992) Identification of a neutralizing domain in the external envelope glycoprotein of simian immunodeficiency virus. AIDS Res Hum Retroviruses 8: 1165–1170.
35. Kent KA, Rud E, Corcoran T, Powell C, Thiriart C, et al. (1992) Identification of two neutralizing and 8 non-neutralizing epitopes on simian immunodeficiency virus envelope using monoclonal antibodies. AIDS Res Hum Retroviruses 8: 1147–1151.
36. Matsumi S, Matsushita S, Yoshimura K, Javaherian K, Takatsuki K (1995) Neutralizing monoclonal antibody against a external envelope glycoprotein (gp110) of SIVmac251. AIDS Res Hum Retroviruses 11: 501–508.
37. Jurkiewicz E, Hunsmann G, Schaffner J, Nisslein T, Luke W, et al. (1997) Identification of the V1 region as a linear neutralizing epitope of the simian immunodeficiency virus SIVmac envelope glycoprotein. J Virol 71: 9475–9481.
38. Pinter A, Honnen WJ, He Y, Gorny MK, Zolla-Pazner S, et al. (2004) The V1/V2 domain of gp120 is a global regulator of the sensitivity of primary human immunodeficiency virus type 1 isolates to neutralization by antibodies commonly induced upon infection. J Virol 78: 5205–5215.
39. Lamers S, Sleasman JW, She JX, Barrie KA, Pomeroy SM, et al. (1993) Independent variation and positive selection in *env* V1-V2 domains within maternal-infant strains of human immunodeficiency virus type-1 in vivo. JVirol 67: 3951–3960.
40. Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, et al. (2004) Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques. J Virol 78: 3561–3571.
41. Frost SD, Wrin T, Smith DM, Pond SL, Liu Y, et al. (2005) Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci U S A 102: 18514–18519.
42. Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, et al. (2006) Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. J Virol 80: 5211–5218.
43. Johnson WE, Morgan J, Reitter J, Puffer BA, Czajak S, et al. (2002) A replication-competent, neutralization-sensitive variant of simian immunodeficiency virus lacking 100 amino acids of envelope. J Virol 76: 2075–2086.
44. Cao J, Sullivan N, Desjardin E, Parolin C, Robinson J, et al. (1997) Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. J Virol 71: 9808–9812.
45. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, et al. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science 303: 2019–2022.
46. Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, et al. (2008) Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. Virology 374: 229–233.
47. Frost SD, Liu Y, Pond SL, Chappey C, Wrin T, et al. (2005) Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. J Virol 79: 6523–6527.
48. Hughes ES, Bell JE, Simmonds P (1997) Investigation of population diversity of human immunodeficiency virus type 1 in vivo by nucleotide sequencing and length polymorphism analysis of the V1/V2 hypervariable region of env. J Gen Virol 78((Pt 11)): 2871–2882.
49. Schacker T, Collier AC, Hughes J, Shea T, Corey L (1996) Clinical and epidemiologic features of primary HIV infection. Ann Intern Med 125: 257–264.

50. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, et al. (1987) The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 126: 310–318.
51. Poss M, Rodrigo AG, Gosink JJ, Learn GH, de Vange Panteleeff D, et al. (1998) Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J Virol 72: 8240–8251.
52. Dacheux L, Moreau A, Ataman-Onal Y, Biron F, Verrier B, et al. (2004) Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. J Virol 78: 12625–12637.
53. Liu Y, McNevin J, Cao J, Zhao H, Genowati I, et al. (2006) Selection on the human immunodeficiency virus type 1 proteome following primary infection. J Virol 80: 9519–9529.
54. Trkola A, Kuhmann SE, Strizki JM, Maxwell E, Ketas T, et al. (2002) HIV-1 escape from a small molecule, CCR5-specific entry inhibitor does not involve CXCR4 use. Proc Natl Acad Sci U S A 99: 395–400.
55. McDonald RA, Mayers DL, Chung RC, Wagner KF, Ratto-Kim S, et al. (1997) Evolution of human immunodeficiency virus type 1 env sequence variation in patients with diverse rates of disease progression and T-cell function. J Virol 71: 1871–1879.
56. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 73: 10489–10502.
57. Tobin NH, Learn GH, Holte SE, Wang Y, Melvin AJ, et al. (2005) Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. J Virol 79: 9625–9634.
58. Rodrigo AG, Goracke PC, Rowhanian K, Mullins JI (1997) Quantitation of target molecules from polymerase chain reaction-based limiting dilution assays. AIDS Res and Hum Retrovir 13: 737–742.
59. Altfeld M, Rosenberg ES, Shankarappa R, Mukherjee JS, Hecht FM, et al. (2001) Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. J Exp Med 193: 169–180.
60. Delwart EL, Herring B, Rodrigo AG, Mullins JI (1995) Genetic Subtyping of Human Immunodeficiency Virus Using a Heteroduplex Mobility Assay. PCR Methods and Applications 4: S202–216.
61. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
62. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, et al. (2003) Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 77: 13376–13388.
63. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R (2007) Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol 25: 1407–1410.
64. Pillai S, Good B, Richman D, Corbeil J (2003) A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses 19: 145–149.
65. Boisvert S, Marchand M, Laviolette F, Corbeil J (2008) HIV-1 coreceptor usage prediction without multiple alignments: an application of string kernels. Retrovirology 5: 110.
66. Busch MP, Satten GA (1997) Time course of viremia and antibody seroconversion following human immunodeficiency virus exposure. Am J Med 102: 117–124; discussion 125–116.
67. Constantine NT, van der Groen G, Belsey EM, Tamashiro H (1994) Sensitivity of HIV-antibody assays determined by seroconversion panels. Aids 8: 1715–1720.
68. Hanley JA, Negassa A, Edwardes MD, Forrester JE (2003) Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 157: 364–375.
69. Burton P, Gurrin L, Sly P (1998) Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. Stat Med 17: 1261–1291.
70. Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42: 121–130.
71. Edmonson P, Murphey-Corb M, Martin LN, Delahunty C, Heeney J, et al. (1998) Evolution of a Simian Immunodeficiency Virus pathogen. J Virol 72: 405–414.
72. Bunnik EM, Pisas L, van Nuenen AC, Schuitemaker H (2008) Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. J Virol 82: 7932–7941.
73. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, et al. (1996) HIV quasispecies and resampling. Science 273: 415–416.
74. Delwart EL, Pan H, Sheppard HW, Wolpert D, Neumann AU, et al. (1997) Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. J Virol 71: 7498–7508.
75. Pilcher CD, Tien HC, Eron JJ, Jr., Vernazza PL, Leu SY, et al. (2004) Brief but Efficient: Acute HIV Infection and the Sexual Transmission of HIV. J Infect Dis 189: 1785–1792.
76. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, et al. (2005) Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis 191: 1403–1409.
77. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, et al. (2009) Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641–645.
78. Moore CB, John M, James IR, Christiansen FT, Witt CS, et al. (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 296: 1439–1443.
79. Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, et al. (2002) Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J Virol 76: 8757–8768.
80. Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, et al. (1995) Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol 69: 8118–8122.