



Identification of amino acid substitutions associated with neutralization phenotype in the human immunodeficiency virus type-1 subtype C gp120

Jennifer L. Kirchherr^{a,b}, Jennifer Hamilton^{a,b}, Xiaozhi Lu^{a,b}, S. Gnanakaran^d, Mark Muldoon^e, Marcus Daniels^d, Webster Kasongo^f, Victor Chalwe^f, Chanda Mulenga^f, Lawrence Mwananyanda^f, Rosemary M. Musonda^f, Xing Yuan^c, David C. Montefiori^c, Bette T. Korber^{d,g}, Barton F. Haynes^{a,b,*}, Feng Gao^{a,b,*}

^a Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC 27710, USA

^b Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA

^c Department of Surgery, Duke University Medical Center, Durham, NC 27710, USA

^d Los Alamos National Laboratory, Los Alamos, NM 87545, USA

^e School of Mathematics, University of Manchester, Manchester, M13 9PL, UK

^f The Tropical Disease Research Centre, PO Box 71769 Ndola, Zambia

^g Santa Fe Institute, Santa Fe, NM 87501, USA

ARTICLE INFO

Article history:

Received 13 July 2010

Returned to author for revision

22 August 2010

Accepted 27 September 2010

Available online 30 October 2010

Keywords:

HIV-1

Envelope

Genetic variation

Neutralization

Signature

ABSTRACT

Neutralizing antibodies (Nabs) are thought to play an important role in prevention and control of HIV-1 infection and should be targeted by an AIDS vaccine. It is critical to understand how HIV-1 induces Nabs by analyzing viral sequences in both tested viruses and sera. Neutralization susceptibility to antibodies in autologous and heterologous plasma was determined for multiple Envs (3–6) from each of 15 subtype-C-infected individuals. Heterologous neutralization was divided into two distinct groups: plasma with strong, cross-reactive neutralization ($n = 9$) and plasma with weak neutralization ($n = 6$). Plasma with cross-reactive heterologous Nabs also more potently neutralized contemporaneous autologous viruses. Analysis of Env sequences in plasma from both groups revealed a three-amino-acid substitution pattern in the V4 region that was associated with greater neutralization potency and breadth. Identification of such potential neutralization signatures may have important implications for the development of HIV-1 vaccines capable of inducing Nabs to subtype C HIV-1.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The ability to elicit cross-reactive neutralizing antibodies (Nabs) is an important goal for HIV-1 vaccines (Burton et al., 2004; Haynes and Montefiori, 2006). HIV-1 has nine genetically related lineages (subtypes A–K), and at a minimum, at least one clade should be effectively targeted in an HIV vaccine for the vaccine to be useful in a part of the world where a single subtype dominates the epidemic (Korber and Gnanakaran, 2009; Korber et al., 2009). Nabs capable of targeting subtype C variants of the virus would be particularly useful, since subtype C accounts for approximately 50% of all HIV-1 infections world-wide (Hemelaar et al., 2006). In addition, the HIV epidemic in large regions of Southern Africa and India is almost completely dominated by subtype C infections; hence in this study, we focus on subtype C antibody responses in natural infection. Efforts to generate

cross-reactive Nabs have had limited success, and novel approaches are urgently needed (Burton et al., 2004; Desrosiers, 2004). Cross-reactive neutralizing activity that is seen with a subset of human monoclonal Abs (Binley et al., 2004; Lin and Nara, 2007; Moore et al., 1994; Trkola et al., 1995; Wu et al., 2010; Zwick et al., 2001) and serum samples from HIV-1-infected individuals (Binley et al., 2008; Dhillon et al., 2007; Li et al., 2007; Sather et al., 2009; Shen et al., 2009; Stamatatos et al., 2009) is evidence that improvements in immunogen design may be possible.

Insights into how to make such improvements are being sought by studying autologous and heterologous Nabs in HIV-1-positive serum samples. Recent results indicate that epitopes in and around the CD4 binding site of gp120 comprise key targets for broadly neutralizing HIV-1-positive sera and that other key targets that play a substantial role remain to be identified (Binley et al., 2008; Dhillon et al., 2007; Li et al., 2007; Wu et al., 2010; Zhou et al., 2010). Studies of autologous virus neutralization have shown that the response targets multiple regions of gp120, most notably epitopes in V1/V2 (Rong et al., 2009) and epitopes that require an interaction between C3 and V4 (Moore et al., 2008, 2009; Rademeyer et al., 2007; Rong et al., 2009). A recent

* Corresponding author. 3072B MSRB II, DUMC 103020, 106 Research Drive, Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC 27710, USA. Fax: +1 919 681 8992.

E-mail address: fgao@duke.edu (F. Gao).

study identified 19 signatures in gp120 and 14 signatures in gp41 associated with neutralization susceptibility of a multi-subtype panel of viruses (Kulkarni et al., 2009). Walker et al. have recently found two mAbs that bind to conformation determinants of HIV-1 Env and broadly neutralize about two-thirds of viruses tested (Walker et al., 2009).

Previous studies of the autologous and heterologous Nab responses in HIV-1 infection have utilized one or a limited number of representative *env* genes from each individual to characterize neutralization susceptibility (Cham et al., 2006; Derdeyn et al., 2004; Wei et al., 2003; Zhang et al., 2007). Since *env* is highly variable in chronic HIV-1 infection (CHI) and because minor sequence changes can affect the biological function and antigenicity of the envelope glycoproteins (Cordonnier et al., 1989; Kalia et al., 2005; LaBranche et al., 1995; Morris et al., 1994; Shimizu et al., 1999; Shioda et al., 1994), the study of a single *env* gene from each infected individual provides only a minimal representation of viral populations *in vivo*. Thus, depending on the design of the study, key information about neutralization epitopes may be missed under these conditions.

Another limitation of previous studies is the reliance on a traditional bulk PCR methodology for Env cloning (Cham et al., 2006; Derdeyn et al., 2004; Palmer et al., 2005; Wei et al., 2003; Zhang et al., 2007). Viral sequences from the quasispecies population obtained by bulk PCR can result in artificial recombination and resampling as well as in nucleoside misincorporation by low-fidelity *Taq* polymerase (Fang et al., 1998; Liu et al., 1996; Salazar-Gonzalez et al., 2008). The single-genome amplification (SGA) methodology makes it possible to obtain *bona fide* viral genomes from the infected individual (Keele et al., 2008; Kirchherr et al., 2007; Palmer et al., 2005; Salazar-Gonzalez et al., 2008). Because viral sequences obtained by SGA more accurately reflect what is present *in vivo*, they can be used to better characterize viral gene functions.

The study of multiple *env* genes in chronic infection in autologous and heterologous neutralization assays enabled us to explore the question of whether there are common neutralization signatures that associate with cross-reactive Nab responses among clade C viral sequences. We have utilized SGA and a novel promoter PCR method to express functional Envs in a high-throughput format (Kirchherr et al., 2007). Multiple *env* genes from each of 37 HIV-1-infected individuals were obtained and characterized with respect to their infectivity and their susceptibility to neutralization by autologous and heterologous plasma samples. After scanning the full Env for potential signature sites, we found potential signature amino acids in the fourth variable region (V4) of gp120 that were associated with cross-reactive Nab responses in subtype C HIV-1-infected individuals; this region has been shown to be critical for NAb susceptibility in the C subtype (Moore et al., 2008).

Results

Genetic analysis of full-length *env* genes

Plasma samples were collected from 50 CHI individuals in Ndola, Zambia, in 2005. A total of 474 complete *env* genes were obtained from 37 of 50 subjects by SGA. Negative PCR results for the remaining 13 subjects were due to either low or undetectable plasma viral loads. An average of thirteen *env* SGAs (range 5–23) were sequenced from each of the 37 PCR-positive subjects (Table 1). Phylogenetic analysis showed that the majority of subjects ($n=34$) were infected with HIV-1 subtype C (Fig. 1). Three others were infected with either subtype D, an A/C recombinant (subtype C in the most central region), or a G/J recombinant.

Identical or closely related *env* sequences were identified in 13 of 37 (35%) of subjects (Table 2), suggesting frequent clonal expansion of minority viral populations in the HIV-1-infected individuals. In these subjects, clonally expanded Envs comprised 9–38% of the

Table 1

Summary of SGAs and functional *env* genes from HIV-1-infected individuals.

ID	Gender	Age	Viral load	Subtype	No. of SGAs	No. of functional <i>env</i> genes	Functional <i>env</i> genes (%)
ZM373	M	25	44,800	C	5	4	80
ZM375	F	31	52,800	C	10	8	80
ZM376	F	30	105,600	C	15	13	87
ZM377	M	41	248,800	C	16	15	94
ZM378	F	36	67,200	C	13	11	85
ZM379	M	35	97,600	C	14	11	79
ZM380	M	37	268,800	A/C	10	10	100
ZM381	F	36	487,200	C	13	12	92
ZM382	M	39	557,600	C	10	10	100
ZM383	F	31	20,240	C	10	5	50
ZM384	M	41	15,440	C	11	10	91
ZM387	M	30	1,120	G/J	9	9	100
ZM388	F	28	21,360	C	11	6	55
ZM389	M	33	1,520,000	C	10	10	100
ZM393	F	21	<384	C	11	2	18
ZM394	F	30	197,600	C	19	12	63
ZM395	F	22	776,800	C	21	20	95
ZM399	M	36	68,240	C	14	14	100
ZM400	M	35	103,200	C	18	15	83
ZM401	M	34	84,000	C	11	9	82
ZM402	F	37	3,920	C	9	6	67
ZM403	F	45	48,080	C	11	10	91
ZM405	F	36	13,920	C	12	9	75
ZM406	M	27	104,000	C	10	10	100
ZM407	M	38	283,200	D	11	9	82
ZM408	F	26	42,400	C	10	8	80
ZM410	M	42	143,200	C	11	6	55
ZM411	F	41	61,600	C	14	6	43
ZM412	F	30	260,000	C	10	7	70
ZM413	F	34	155,200	C	16	14	88
ZM414	F	25	213,600	C	22	14	64
ZM415	M	38	64,720	C	21	18	86
ZM416	M	45	80,800	C	23	23	100
ZM417	M	32	96,000	C	9	9	100
ZM418	F	21	532,000	C	10	8	80
ZM419	F	33	433,600	C	10	1	10
ZM420	F	24	286,400	C	14	13	93
Total					474	377	80

sampled *env* population. Clonal expansion sequences formed a tight cluster with zero or very short branch lengths in the phylogenetic trees (Fig. 2); this tight clustering was supported by high bootstrap values (>85%). Nine subjects had one clonally expanded *env* variant (Fig. 2A–I), whereas another four subjects (ZM377, ZM405, ZM414, and ZM415) had two independent clonally expanded populations (Fig. 2J–M). Clonally expanded variants in two subjects (ZM402 and ZM405) formed new lineages (Fig. 2E and K). Sequences closely related to, but beginning to diverge from the clonally expanded variants, were observed in four subjects (ZM377, ZM401, ZM414, and ZM415) (Fig. 2D, J, L, and M). Recombinant *env* sequences were also detected between clonally expanded viruses and other viruses in ZM394 (Fig. 2I). These latter recombinants contributed to a substantial increase in diversity of the viral population in the subject. When recombinant sequences were excluded from the phylogenetic tree, the clonally expanded viruses represented an emergence of a new viral lineage in the subject (Fig. 2I).

Functional analysis of the *env* quasispecies in infected individuals

A pPCR method that adds a CMV promoter to SGA amplicons was used for high-throughput functional screening of multiple *env* genes from each infected individual (Kirchherr et al., 2007). Env-pseudotyped viruses were generated by cotransfecting 293T cells with pPCR products and an Env-defective backbone plasmid (SG3Δenv).

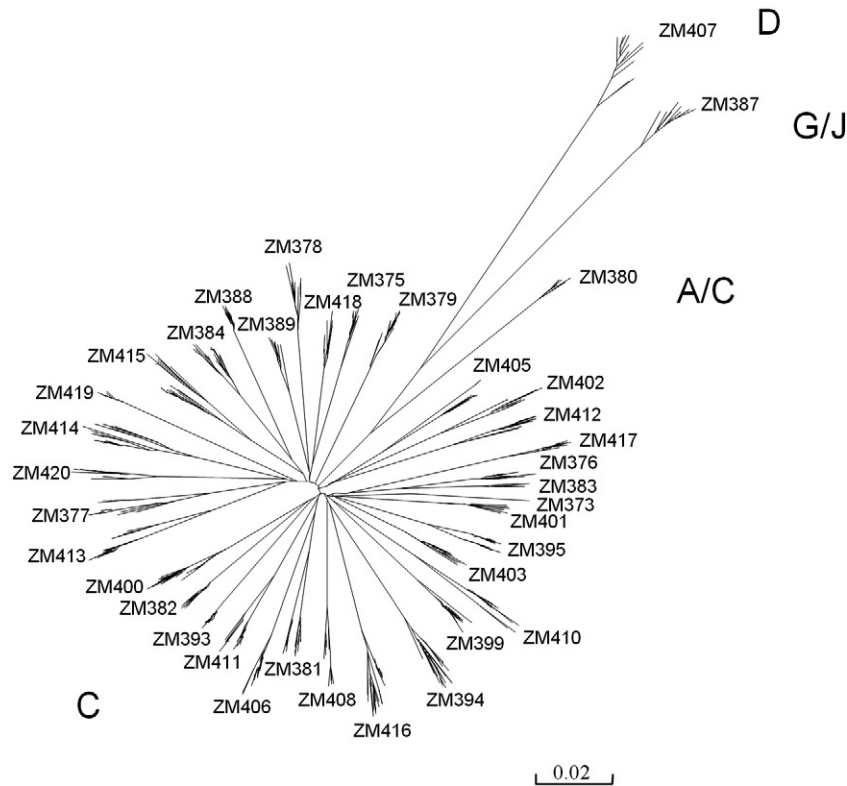


Fig. 1. Phylogenetic tree analysis of newly characterized full-length *env* gene sequences. An unrooted phylogenetic tree was constructed with complete *env* gene sequences using the neighbor-joining method and the Kimura two-parameter model. The branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site).

Infectivity was measured by luciferase activity in TZM-bl cells. Of the 474 Envs examined, 377 (80%) were found to be functional (Table 1). All *env* genes were functional in eight subjects, whereas in a few subjects, only a small fraction of *env* genes were functional (10% and 18% for ZM419 and ZM393, respectively). In three individuals (ZM373, ZM380, and ZM382), all Env-pseudotyped viruses were highly infectious and differed in infectivity by only a few fold (Fig. 3). Notably, infectivity of Env-pseudotyped viruses from most other individuals was highly variable (1–2 log differences in luciferase activity).

Recombinant genes were found in 11 *env* sequences from 6 individuals infected with two lineage viruses. Examination of their functionality showed that 8 (73%) were biologically functional while the other 3 were not. ZM420.9 was not functional due to a reading frame shift caused by one nucleotide insertion. The result suggested that while some recombinants might gain biological advantages, others might be lethal and result in nonfunctional genes. For example, in ZM394, the recombinant viruses between the clonal expansion viruses and other viruses resulted in different levels of infectivity, ranging from nonfunctional to highly infectious (Figs. 2I and 3). However, recombinant sequences were not particularly enriched for inactive viruses relative to non-recombinant (Fisher's Exact Test, $p=0.4$).

Given the range of variability in Env function, it was of interest to determine whether the level of Env infectivity in pseudovirus assays was associated with either protein expression levels or with plasma viral RNA load. As shown in Fig. 4, a trend towards a positive correlation was seen between pseudovirus infectivity and viral load ($p=0.05$). Protein expression in transfected 293T cells was quantified by Western blot for multiple *env* SGAs from three subjects (ZM400, ZM413, and ZM414) whose Env clones demonstrated a substantial range in infectivity. To control for variation between independent

transfections, Env expression was standardized to Gag protein P24 expression in the same cells. The sizes and processing cleavage efficiency of gp160 into gp120 and gp41 were different among tested *env* genes. While all Envs showed little size variation in ZM414 (except the prematurely truncated ones), they varied in both ZM400 and ZM413 (Fig. 5). The cleaved gp120 was observed for nearly all expressed Envs in ZM400, but most were uncleaved gp160 in ZM413 and ZM414. Infectivity of Env pseudoviruses did not correlate with the level of Env expression in each individual.

These results suggested that variation in infectivity was not determined by expression levels or the cleavage efficiency of gp160 in transfected cells. It is possible that the amount of Env in transfected cells was above the threshold required for the production of infectious pseudoviruses since only a small number of Env spikes are present on mature HIV-1 particles (Chertova et al., 2002; Zhu et al., 2003). Non-infectious pseudoviruses were associated with either the absence of Env expression or with the expression of truncated Envs; these cases were explained by either premature stop codons or frame-shifting deletions by sequence analysis.

Autologous and heterologous neutralization

A checkerboard-style dataset of the neutralizing activity in autologous and heterologous plasma samples was generated by assaying plasma samples against multiple representative Env-pseudotyped viruses from each subject. Because of the limited supply of plasma, it was only possible to perform neutralization assays with 60 Env-pseudotyped viruses from 15 subjects (14 subtype C and 1 A/C recombinant) using plasma samples from these 15 subjects plus one additional plasma sample (ZM383). Multiple *env* genes (range 3–6) from each subject were selected to represent the viral population as seen in a Maximum likelihood tree analysis (Fig. 6). Three or four *env*

Table 2Characterization of clonal expansion *env* sequences in HIV-1-infected individuals.

ID	SGA No.	Luciferase activity (RLU)	Total viral population (%)	No. of amino acid differences among <i>env</i> sequences
ZM375	1	434,052	30	3, 4, 5
	9	437,851		
	11	610,112		
ZM377	10	175,205	19	1, 2, 3
	13	310,710		
	14	53,926		
	11	354,745		
ZM378	12	30,765	19	1, 4, 5
	16	77,427		
	7	802,968		
ZM394	9	50,041	15	3
	11	398,245		
ZM395	18	21,273	10	2
	14	327,566		
ZM401	15	192,620	10	1
	2	164,195		
ZM402	19	66,826	27	2
	20	161,349		
	8	48,981		
ZM405	12	8,985	20	0
	25	205,738		
ZM408	44 ^a	1,393	17	4
	43	92,501		
	53	118,441		
ZM411	15	118,979	20	3
	23 ^a	2,218		
ZM413	6	92,343	14	3
	9	203,082		
ZM414	5	229,782	13	3
	13	348,872		
ZM415	1	47,219	18	0, 1
	10	70,754		
	23	218,271		
	28	4,134		
ZM416	9	141,022	14	1
	20	240,243		
	25	126,764		
	1	510,266		
ZM417	2	114,320	16	4
	26	553,315		
	15	496,933		
	27	297,937		
ZM418	3	391,763	9	2
	16	296,184		
Control	SG3Δenv	1,127		

^a Stop codon.

genes were chosen for monophyletic viruses, whereas 1–3 *env* genes were selected from each lineage in individuals with multiple *Env* lineages.

Neutralization results are summarized in Supplementary Table 1. A majority of virus/plasma combinations (75%) were positive with neutralization titers of 20–3769, where approximately one-third of the combinations resulted in a neutralization titer >100. Variable levels of autologous virus neutralization by contemporaneous plasma (i.e., plasma obtained at that same time point as *Env*) were detected in all but one subject (ZM380). The neutralization potency was relatively weak in most cases; however, titers >100 were detected against at least one variant in the autologous virus quasispecies of 7 subjects (ZM378, ZM379, ZM401, ZM408, ZM413, ZM416, and ZM417). In several cases, substantial differences were seen between multiple *env* genes from the same subject. For example, titers of autologous neutralization against multiple *env* variants ranged from 24 to 598 in subject ZM408 and from 37 to 449 in subject ZM416.

All *Env*-pseudotyped viruses were then assayed for neutralization susceptibility by heterologous plasma. Plasma samples were segregated into low (L) and high (H) neutralization potency groups by heatmap analysis (Fig. 7). Statistically supported clustering of plasma samples was observed, with the distinctive low-activity cluster boxed on the left. Overall, seven plasma samples had limited cross-neutralizing activity. However, the other nine plasma samples possessed moderate to potent cross-reactive neutralization activity (Fig. 7 and Supplementary Table 1). Some plasmas (ZM408, ZM378, ZM395, and ZM379) neutralized nearly all 60 pseudoviruses from the 15 infected individuals.

Neutralization signature analysis

Identification of plasma samples with potent and cross-reactive Nabs prompted us to ask if there were signature sequences present among the *Envs* associated with the potent neutralizing activity. Because each plasma sample represented one value per individual, we used only the consensus sequence of all variants within an individual for the signature study, thereby capturing the most common amino acid in each position. Based on the two distinct clusters shown in the heatmap (Fig. 7), we scanned each position in the alignment for amino acids associated with either strong or weak subtype C cross-reactive neutralization responses using the Maximum likelihood tree correction method previously described (Bhattacharya et al., 2007). Because of multiple test issues (a test was done for every site in *Env*), we used a *q*-value to estimate the false discovery rate, based on the distribution of *p*-values (Storey and Tibshirani, 2003).

Four signature sites were identified: three in gp120 and one in gp41 (Fig. 8). Each site had a *p*-value of 0.03, but a *q*-value of 0.27; thus it is probable that one or more of the signatures identified were false positives. Nonetheless, we include them here in a hypothesis-raising mode because they displayed the strongest trends towards significance. Intriguingly, three of the four signature sites were found in the V4 region (393G, 397G, and 413N). The V4 region, in association with the C3 alpha-2 helical domain, is thought to contribute to patterns of neutralization susceptibility in subtype C viruses (Gnanakaran et al., 2007; Moore et al., 2008; Rong et al., 2007) (Fig. 9). The fourth signature was found at the very end of the gp41 cytoplasm domain and consisted of a lysine (L) in position 856. The most intriguing signature was position 413 N, which is part of a potential N-linked glycosylation (PNLG) site signature that was independently identified to be associated with good serological Nab breadth in a multi-subtype study of envelopes isolated from individuals who had elicited particularly potent neutralizing antibodies versus those that had weak responses (Gnanakaran et al., 2010). Thus, *Envs* associated with more cross-reactive Nabs tended to have mutated towards an Asn (N) in position 413. They also tended to have mutations towards Gly (G) in positions 393 and 397, although position 397 was in a highly variable region that is difficult to align with confidence. Interestingly, autologous neutralizing activity was significantly higher (*p*=0.0016) in the high neutralization plasma group than in the low-neutralization plasma group of heterologous neutralization potencies (Fig. 10). In other words, plasmas that possessed stronger neutralizing activity against heterologous viruses more potently neutralized the contemporaneous autologous virus. In addition, positions 393, 397, and 413 were found to be highly variable within single individuals, indicative of recurrent immune pressure in different infections at these sites. Collectively, these data suggest that signature patterns in the V4 loop are associated with potent and cross-reactive neutralizing antibodies against subtype C viruses.

Discussion

We characterized multiple HIV-1 *env* genes in each of 37 individuals using SGA and pPCR methods. Sequence and functional

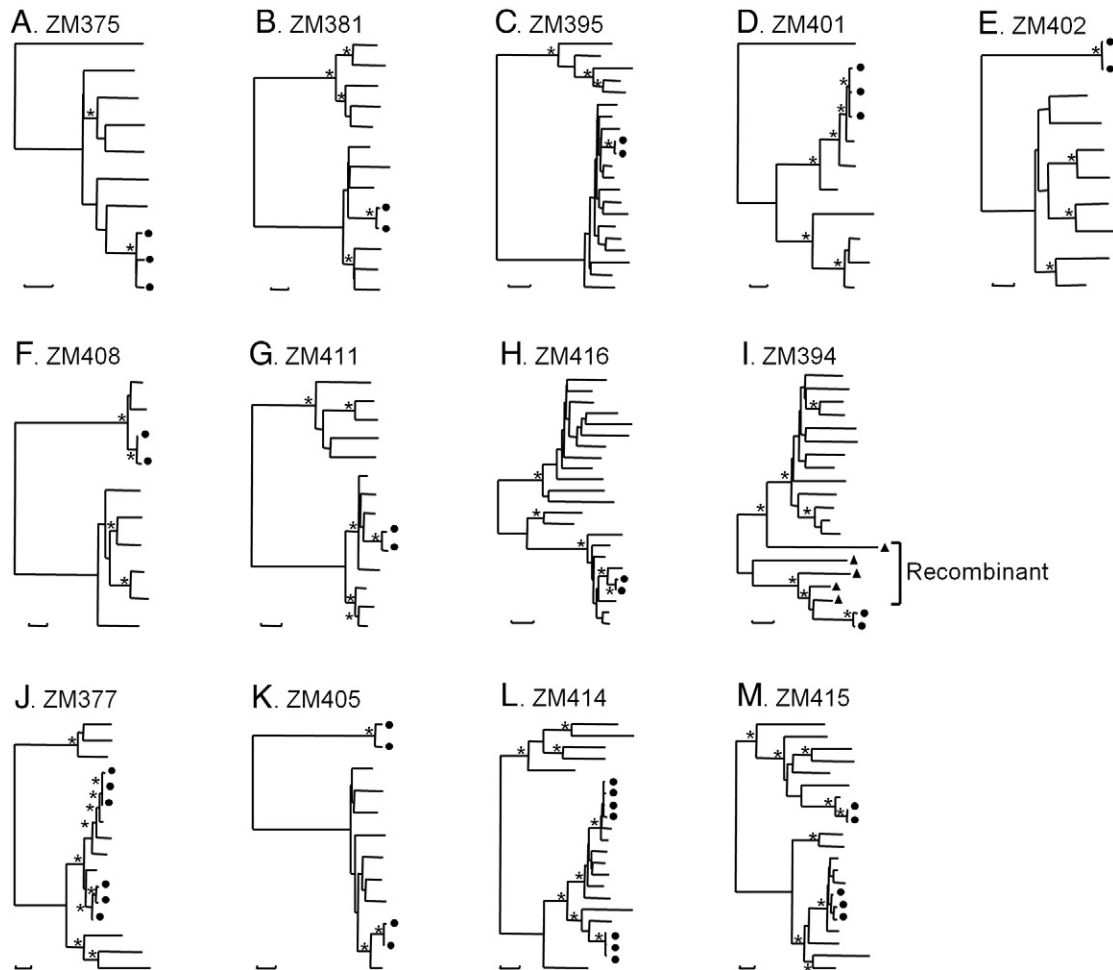


Fig. 2. Phylogenetic tree analysis of clonal expansion *env* gene sequences. Midpoint-rooted phylogenetic trees were constructed with all *env* sequences from each individual who harbored clonal expansion viruses using the neighbor-joining method and the Kimura two-parameter model. Horizontal branch lengths are drawn to scale (the scale bar represents 0.005 nucleotide substitutions per site); vertical separation is for clarity only. Asterisks indicate bootstrap values in which the cluster to the right is supported in 85% or more replicates (out of 1000). Sequences with identical or nearly identical sequences are marked with black dots at the tip of the tree branches.

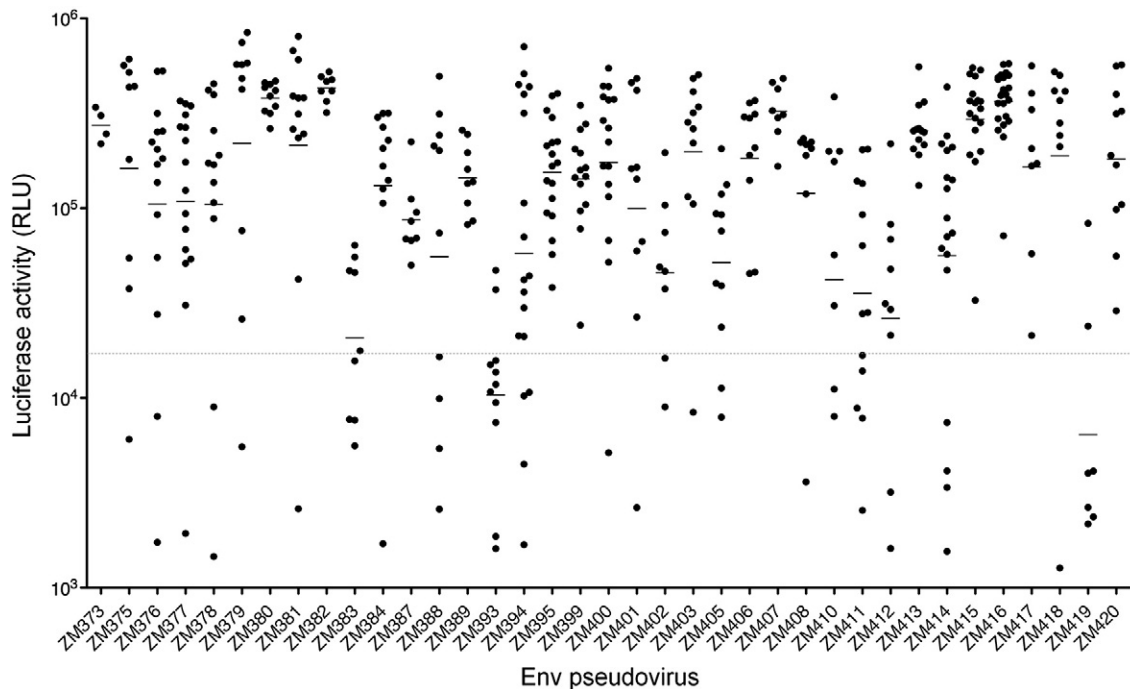


Fig. 3. Infectivity of Env pseudoviruses. The infectivity of 474 pseudoviruses from 37 individuals was determined in TZM-bl cells. The *env* genes were considered positive when the luciferase activity RLU values were 10-fold greater than those from SG3Δ*env* backbone control. The dotted line indicates this cutoff.

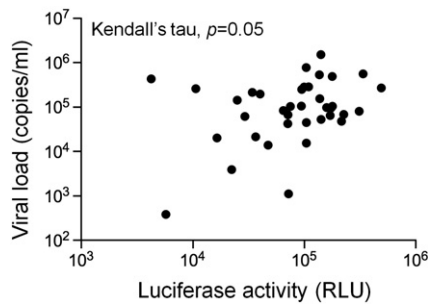


Fig. 4. Correlation between viral load and infectivity of Env pseudoviruses. Geometric means of the average luciferase activity of multiple Env pseudoviruses from each HIV-1-infected individual were plotted and analyzed by the Kendall's tau rank correlation coefficient method.

analysis of the *env* gene quasispecies in each individual showed frequent clonal expansions, a trend towards a positive correlation between viral loads and infectivity of Env pseudoviruses, highly variable infectivity among *env* gene quasispecies populations, and, importantly, signature amino acids in the V4 region of sequences in

plasma with high levels of cross-reactive HIV-1 neutralization activity. Identification of such neutralization signatures may have implications for the development of effective HIV-1 vaccines.

As expected for chronic HIV-1 infection, a highly divergent viral quasispecies population was detected in the subjects in this study. Intra-subject *env* gene diversity was as high as 8%. Using our recently established pPCR method, we were able to characterize the phenotypic properties of a large number of *env* genes from each HIV-1-infected individual by generating pseudoviruses without cloning SGA PCR products. Analysis of the infectivity of 474 Env pseudoviruses from 37 individuals revealed that a high percentage (80%) of the *env* genes were functional. This is consistent with our previous report with a smaller number of *env* genes (Kirchherr et al., 2007), but much higher than the rate (10%) from *env* gene populations characterized by bulk PCR from PBMC (Cham et al., 2006). The different rates were likely due to differences in methods and sample sources used to obtain *env* genes. It is likely that nonfunctional PCR-generated recombinants were avoided by SGA (Keele et al., 2008; Kirchherr et al., 2007; Palmer et al., 2005; Salazar-Gonzalez et al., 2008).

Functional analysis showed that the infectivity of Env pseudoviruses from the same individuals varied significantly (1–2 logs) in

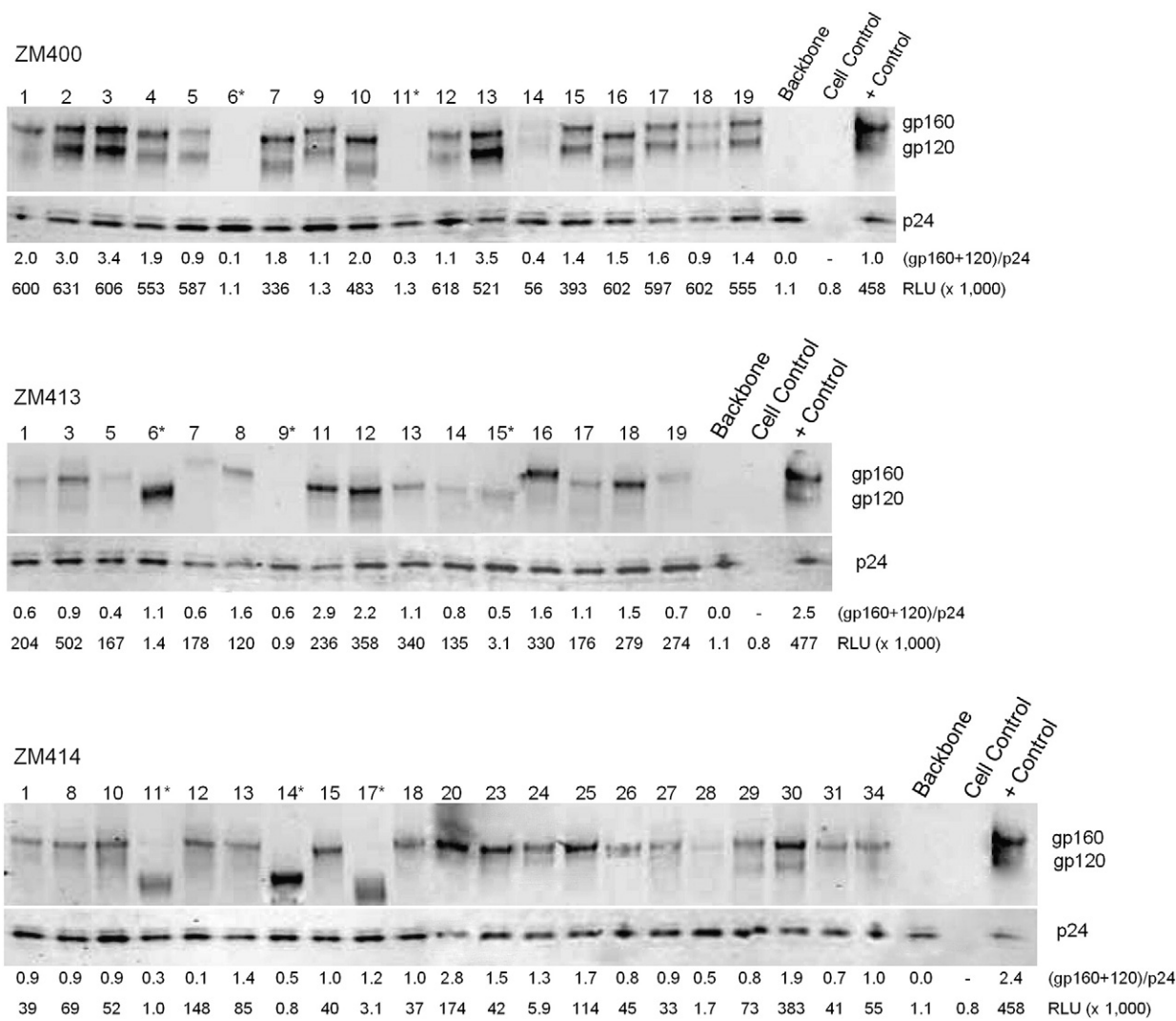


Fig. 5. Western blot analysis of HIV-1 protein expression in transfected cells. The 293T cells transfected with pPCR products and pSG3Δ*env* were lysed 48 h after transfection. The viral proteins were separated on a 4–12% gradient reducing gel, were transferred to nitrocellulose, and were reacted with an HIV-1-positive serum and a mouse mAb 13D5 (Gao et al., 2009) to the Env protein. Viral proteins were visualized with secondary antibodies IRDye800-conjugated goat anti-human and Alexa-Fluor680 goat anti-mouse using a LiCor Odyssey Infrared Imaging system. The expression levels of Env proteins are expressed as Env (gp160 + gp120):P24 ratios. The infectivity of the each Env pseudovirus is shown as relative light unit (RLU). Asterisks indicate the smaller Envs due to premature stop codons or non-inframe deletions.

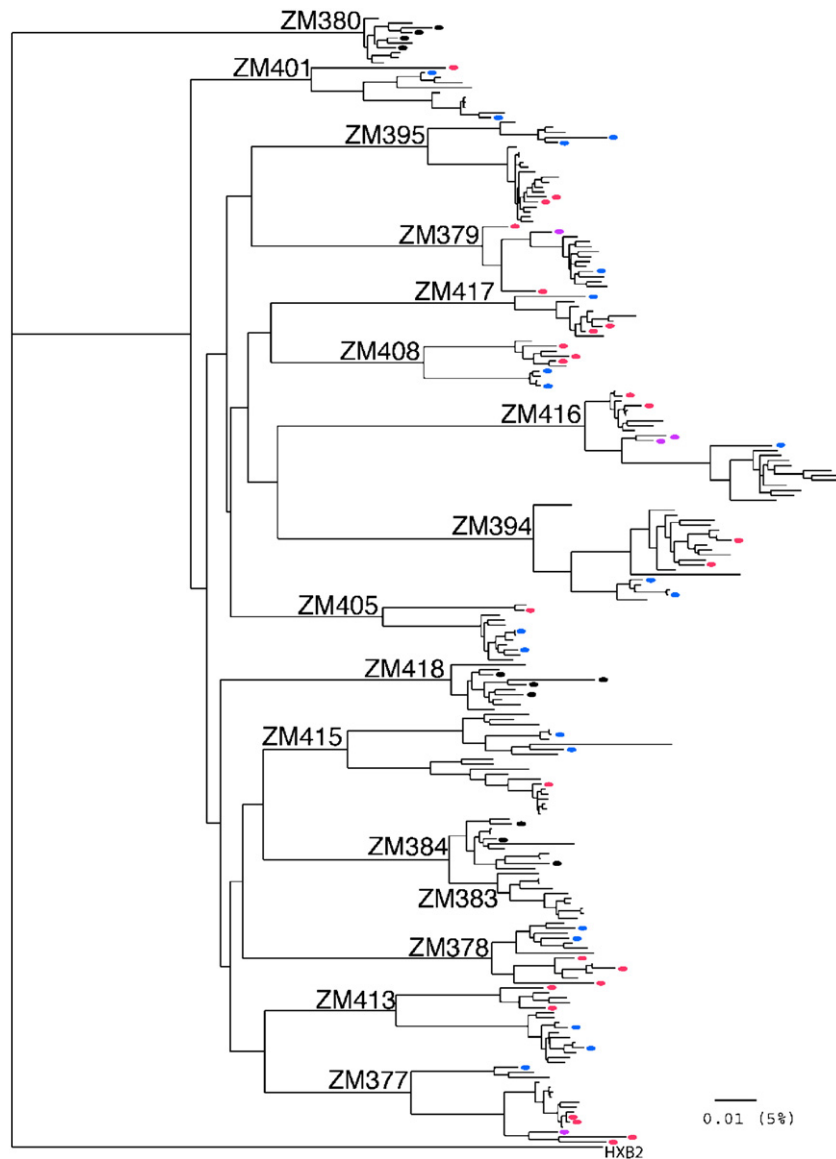


Fig. 6. Maximum likelihood phylogenetic tree of the *env* sequences. All SGA *env* sequences from 15 HIV-1-infected individuals were included for analysis. Those used for neutralization assays are indicated with colored marks at the tip of the tree branches. Envelopes were selected to be representative of the diversity in the sample. The sequences from samples with low diversity (monophyletic) are indicated in black; the sequences from samples with two distinctive phylogenetic clusters are colored so that those from one cluster are red and those from the other are blue; and sequences that represent recombinants between the two clusters are purple.

most individuals. This is similar to plasma-derived *env* genes from one individual (Huang et al., 2008) but different from PBMC-derived *env* genes, which showed uniform infectivity from the same individual (Cham et al., 2006). Examination of Env proteins in the transfected cells showed that the variation of infectivity among Env pseudoviruses in each individual was not correlated with *in vitro* expression levels. With an intact *env* open-reading frame, the nonfunctionality of the *env* genes may be caused by mutations that affect either incorporation into virus particles or poor interaction with CD4 and coreceptors on target cells. Interestingly, a positive correlation trend was observed between the levels of infectivity of Env pseudoviruses and viral loads in the donor subjects, suggesting that viruses with more infectious *env* genes may lead to a high level of viral loads *in vivo*.

Sequence analysis showed that 13 of 37 individuals (35%) carried one or two clusters of closely related sequences among the quasispecies viral populations. These sequences, unique to each individual and conserved within that individual, were identical or

nearly identical to each other. Since each sequence was obtained from independent PCR by SGA, they represent *in vivo* derived independent viral genomes. The *env* sequences related to but beginning to diverge from the identical or nearly identical sequences, as well as recombinants between these conserved sets and more diverse viruses from that individual, indicate that there is clonal expansion of some viruses in the infected individuals. They generally accounted for 9–38% of viral population. Clonal expansion of some virus species was found in different tissues in HIV-1-infected individuals (Bull et al., 2009). In another study, clonal expansion viruses and other divergent viruses were found in an infected mother a day after delivery, but only viruses identical or highly similar to those clonal expansion viruses were detected in the infected infant at birth (Verhofstede et al., 2003). The result from this study strongly suggests that clonal expansion viruses may play an important role in virus transmission and pathogenesis. The current study cannot explain the sources for clonal expansion viruses in plasma. It is possible that they are more fit, selected by immune responses, or from a reservoir activated by

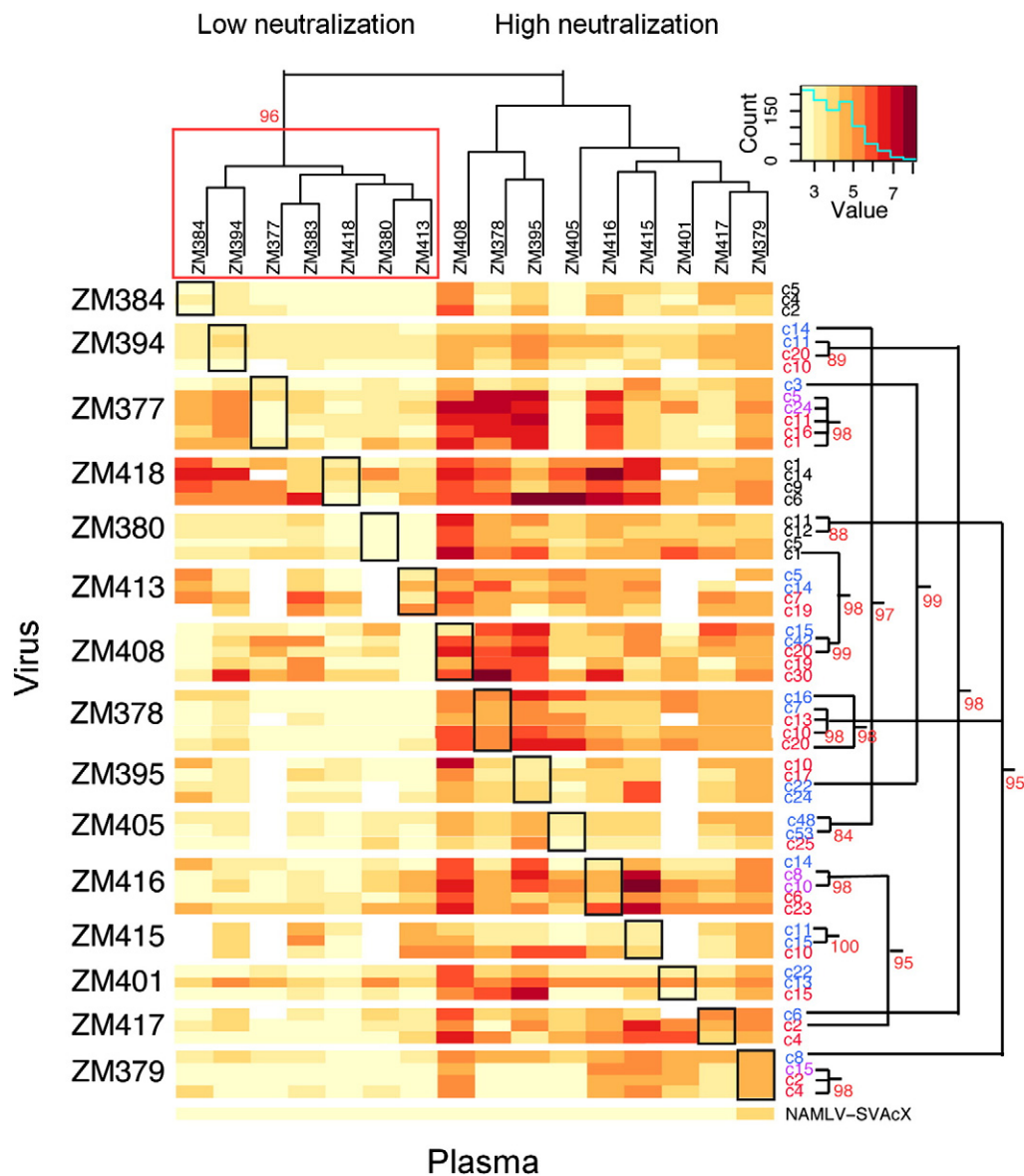


Fig. 7. Hierarchical clustering of viruses and sera based on neutralization titers. Sera are clustered according to their ability to neutralize. Autologous responses are indicated in black boxes along the diagonal. The *env* SGA numbers are colored according to diversity: low diversity (black), two distinct groups (red and blue), and recombinant (purple). For ease of visualization, the heatmap is organized such that the rows (Env pseudoviruses) are arranged according to individuals rather than hierarchical clustering, where hierarchical clustering patterns of the Env pseudoviruses are shown as a dendrogram to the right of the figure. Clustering of high and low-neutralization plasma was statistically supported, with a probability of 0.96 that the distinctive low-neutralization cluster was robust, using the approximately unbiased multistep-multiscale bootstrap resampling method developed by Shimodaira (2004). To illustrate this grouping, the columns are presented according to like-responses to the pseudoviruses based on the clustering hierarchy shown at the top of the figure. The sets of plasma with low (L) or high (H) Nab titers were grouped. Clustering patterns on the right represent significant unbiased multistep-multiscale bootstrap resampling values 95 or greater.

unknown factors. Since they are genetically different from the viruses in the major viral population and can recombine with other viruses (Fig. 2I), they might play a role in evasion of immune responses elicited by the majority viruses in the population. A longitudinal study of multiple time point samples will be required to test this hypothesis.

An analysis of checkerboard neutralization data with multiple representative *env* genes from each individual using autologous and heterologous plasma samples enable us to classify neutralization responses and revealed a signature of four amino acids associated with potent neutralizing plasmas, three of which were near the base of the V4 loop. Amino acids found in signature sites are not expected to be perfectly predictive of a complex state like neutralization phenotype, but the fact that they are enriched in association with such a phenotype indicates that they are important and likely to be

relevant to the state. Glycines associated with signature sites 393 and 397 in the V4 loop of gp120 could impact antibody neutralization in several ways. Firstly, the loss of electrostatic charge associated with side chains can perturb antibody binding by modulating the electrostatic potential of the cognate epitope. Secondly, due to their lack of side chains, glycines exhibit higher flexibility that might affect conformational motions in the V4 loop. Flexibility and less charge repulsion due to glycines at positions 393 and 397 might allow V4 to sample a wide variety of conformations. In contrast, the signature 413N might play a role by mediating N-linked glycosylation and potentially inhibiting V4 loop movement by steric hindrance.

Site 413 is proximal to the 17b and CCR5 binding sites and is adjacent to the proximal binding motif RIKQ (HXB2 residues 419–422). A recent glycopeptide mass spectroscopy (MS) study found that

V4		3	3	4	gp41		8
		9	9	1			5
		3	7	3			6
CONSENSUS-H	GEFFYCNTsdLfn	gt	ngt	nstnsn	ItlpCrIKQIINMWQ	CONSENSUS-H	RIRQGfEaAL
H_ZM378.CONSENSUS	-----K	GEF	-G-R	NNP	NGTSDN	-----Q	-----L
H_ZM379.CONSENSUS	-----G	-K	-----D	-----DKN	-----	-----T	-----L
H_ZM395.CONSENSUS	-----S	SKL	FN.....TN	-----TDNLT	-----I	-----Q	-----Q
H_ZM401.CONSENSUS	-----R	-----G	KF-G	YNR	-----SGN	-----K	-----Q
H_ZM405.CONSENSUS	-----G	-----G	-----G	TDSN	-----ST	-----I	-----L
H_ZM408.CONSENSUS	-----K	-----S	-----G	-----V	IMYDEN	-----ST	-----IQ
H_ZM415.CONSENSUS	-----G	-----I	-----M	PNST...FIS	-----N	-----NKK	-----Q
H_ZM416.CONSENSUS	-----T	-----T	-----M	-----F	E	-----KEN	-----L
H_ZM417.CONSENSUS	-----Q	-----G	-----F	-----G	-----FTRNFTNTSSN	-----LT	-----L
L_ZM384.CONSENSUS	-----TN	-----N	-----RLSE	F	-----S-E	-----ST	-----Q
L_ZM394.CONSENSUS	-----N	-----R	-----SL	-----NDT	EEDTD	-----RT	-----Q
L_ZM377.CONSENSUS	-----K	-----N	-----GT	NMTDP	-----Q	-----	-----Q
L_ZM380.CONSENSUS	-----G	-----S	-----WIFENG	TANSTWP	-----GT	-----Q	-----Q
L_ZM383.CONSENSUS	-----TN	-----T	-----LLDNFI	STG	-----S-G	-----ST	-----Q
L_ZM413.CONSENSUS	-----K	-----S	-----D	NATAENATGT	-----	-----	-----Q
L_ZM418.CONSENSUS	-----G	-----R	-----PNVTNLA	NGTKS	-----NETI	-----R	-----Q

Fig. 8. The signature amino acid sequences associated with high levels of neutralization activity in plasma. H represents sequences in the group of plasma samples with high neutralizing activity against heterologous viruses, whereas L represents sequences that exhibited weak neutralizing activity. Numbers are used to show corresponding locations in the HXB2 reference strain. Dashes indicate the identical amino acids present in the reference sequence. Periods are used to designate gaps to maintain the alignment. Signature sites associated with potent neutralization are shown in red. Non-signature amino acids in key positions are shown in blue.

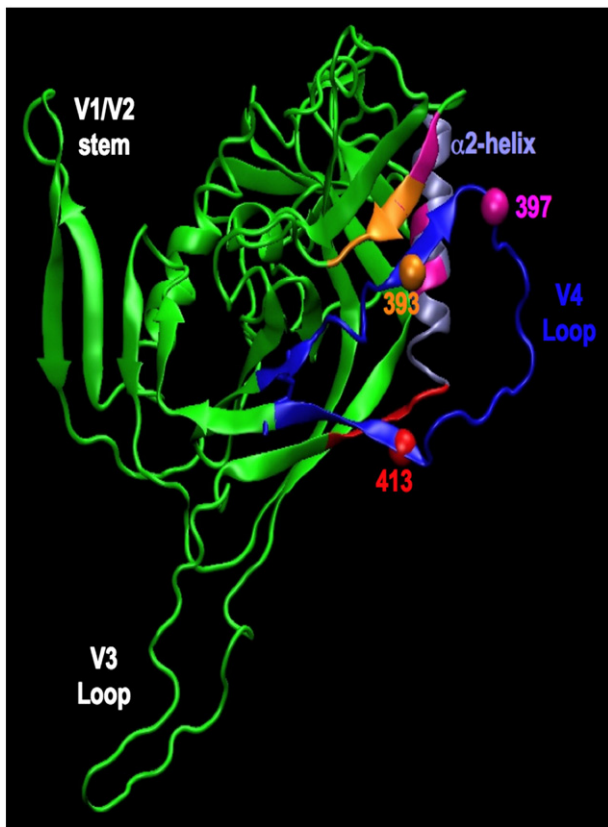


Fig. 9. Crystal structure of liganded gp120 with signature amino acids. The figure shows three signature sites, 393 (orange), 397 (magenta), and 413 (red) on a crystal structure of liganded gp120 (PDB:2B4C) (Huang et al., 2005). The V4 loop and alpha-2 helix are marked in different colors for clarification. Positions in C3 that exhibit significant contact with these signature sites on V4 are also marked with the same color as the signature sites. Three-dimensional images were generated using VMD (Humphrey et al., 1996). Location of marked sites shown on this B clade template for visualization does not show any significant differences if the recently determined X-ray C clade gp120 from CAP210 (Diskin et al., 2010) is used as a template. This is expected since rigorous evaluation of root-mean-square deviation (RMSD) between the CD4-bound truncated gp120 structures from B and C clades showed that both structures are similar (RMSD ~ 1.0, S. Gnanakaran, unpublished analysis).

glycans were indeed attached at this site in the group M consensus HIV-1 envelope, CON-S gp140 (Irungu et al., 2008). Both high-resolution methods, online high-performance liquid chromatography-electrospray ionization mass spectrometry (HPLC/ESI-MS), and offline HPLC followed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) determined that the potential glycosylation site 413 is, in fact, glycosylated. It is not clear, however, how glycosylation at this position is associated with higher levels of neutralizing antibody titers *in vivo*. This position is quite variable; indeed the glycosylation site could be enriched in Envs associated with potent neutralization activity because it is selected against by a strong response rather than being required to elicit a strong response.

In a recent study, a neutralization signature was investigated in subtype C sera using 36 sera against a panel of 5 subtype B and 5 subtype C viruses (Rademeyer et al., 2007). No clear clustering of serological patterns was found to enable tracking of the signature patterns. However, a correlation between the V1–V4 loop length and the cross-subtype neutralizing potential was observed. In the current study, we did not find evidence for such a correlation when comparing the total V1–V4 lengths or the lengths of V1, V2, and V4 separately of the consensus Env from each individual grouped by the high- and low-neutralization plasma using a Wilcoxon rank statistic. However, it could simply be that the number of the samples was

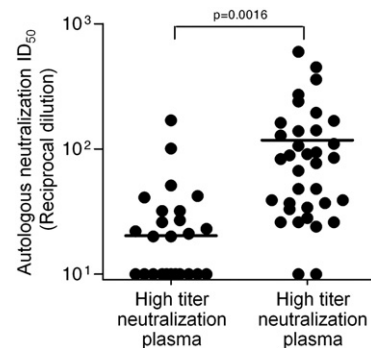


Fig. 10. Comparison of autologous neutralizing activity between low and high heterologous neutralizing plasma samples. Nab titers from low heterologous neutralization plasmas ($n=6$) or high heterologous neutralization plasmas ($n=9$) were compared. Values at Y-axis are the reciprocal plasma dilutions at which luciferase activity (RLU) was reduced by 50% relative to virus control wells by autologous plasma.

underpowered to show this effect. The other possibility is that all samples in the current study were from CHI individuals in whom the length of variable loops had already become longer during the evolution of immune evasion.

Previous studies also showed that the alpha-2 helix was under strong selection pressure (Gnanakaran et al., 2007) and this region was associated with resistance to autologous neutralization of subtype C viruses (Rong et al., 2007). The V4 region is downstream from the alpha-2 helix, and they are structurally close to each other. It was also found that the C3V4 region is likely responsible for inducing autologous and heterologous Nabs (Moore et al., 2008). The newly identified three signature amino acids were located in the vicinity of the previously reported region targeted by Nabs. These signatures may be responsible for eliciting broader and more potent Nab responses or represent the escape mutations from the broadly Nabs against subtype C viruses through similar pathways. In either case, the data suggest that signature patterns in the V4 loop are important to consider for the design of a vaccine that can induce potent and broadly reactive Nabs, at least for subtype C, which is the most prevalent HIV-1 subtype in the world.

Methods and materials

Amplification of HIV-1 *env* genes

Plasma samples were collected from 37 HIV-1-positive individuals enrolled in a study of contemporary HIV-1 strains in Ndola, Zambia. The study was approved by the ethics committee of the Tropical Disease Research Centre, the Duke University Institutional Review Board, and the National Institutes of Health. No people in this study were treated with antiretroviral drugs. The homosexual activity was not reported in the study population. The studied individuals in our study were most likely infected through heterosexual transmission. Viral RNA was extracted from the plasma and reverse transcribed into cDNA using Superscript III (Invitrogen; Carlsbad, CA). Multiple *rev/env* genes from each individual were obtained by using single-genome amplification (SGA), followed by the addition of a CMV promoter to the 5' end of the SGA products using pPCR technology as previously described (Kirchherr et al., 2007).

Single-round infection assay

pPCR products were cotransfected with an *env*-deficient HIV-1 backbone plasmid (pSG3Δ*env*) into 293 T cells in a 24-well plate using FuGENE6 transfection reagent (Roche Diagnostics; Indianapolis, IN). Briefly, pPCR DNA (150 ng) and pSG3Δ*env* DNA (150 ng) were mixed with 1.2 μl of FuGENE6 (FuGENE:DNA ratio at 3 μl:1 μg) in a total volume of 20 μl with serum-free DMEM, incubated for 30 min and added to 293 T cells (70% confluence) seeded 1 day earlier at 5×10^4 per well. Forty-eight hours after transfection, supernatants were harvested. Equal volumes of pseudovirions were added to TZM-bl cells with DEAE (5 μg/ml) in a 96-well plate (200 μl). Cultures were incubated for 48 h at 37 °C with 5% CO₂. Supernatants (100 μl) from infected TZM-bl cells were removed and 100 μl of Bright-Glo Luciferase Assay substrate with buffer (Promega; Madison, WI) was added to the cells. Following a 2-min incubation, 100 μl of cell lysates was added to a solid black 96-well plate. Luminescence was measured with a Wallac 1420 Multilabel Counter (PerkinElmer; Waltham, MA).

Neutralization assay

HIV-1 neutralization was measured as a reduction in luciferase activity after a single-round infection of TZM-bl cells, as previously described (Li et al., 2005, 2006). Equal amounts of pseudovirions (200 TCID₅₀) were used in each reaction. Neutralization titers against pseudoviruses were determined for 16 plasma samples (15 autolo-

gous and 1 heterologous to the tested Env pseudoviruses) and one HIV-1-positive serum control. An amphotropic murine leukemia Env pseudovirus was also included as non-specific neutralization control.

Western blot

Forty-eight hours after cotransfection with pPCR products and pSG3Δ*env*, 293 T cells were lysed with 250 μl of lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 20 mM EDTA, 1% Triton-X100, 0.1% SDS, pH 7.4). Cell lysates were mixed with 6× denaturing sample buffer (300 mM Tris-HCl pH 6.8, 12 mM EDTA, 12% SDS, 864 mM 2-mercaptoethanol, 60% glycerol, 0.05% bromophenol blue). Samples were boiled for 10 min and then loaded on a NuPAGE Novex 4–12% Bis-Tris gel (Invitrogen; Carlsbad, CA). After the samples were transferred to a nitrocellulose membrane, the membrane was blocked in PBS containing 1% casein and 0.01% NaN₃ for 1 h. The blot was reacted with an HIV-1-positive serum (1:500) and a mouse mAb 13D5 (1 μg/ml) to the HIV-1 Env protein. Finally, the membrane was reacted with an IRDye800-conjugated affinity purified goat anti-human IgG antibody (Rockland Immunochemicals; Gilbertsville, PA) and an Alexa-Fluor 680-conjugated goat anti-mouse IgG antibody (Invitrogen; Carlsbad, CA). Fluorescence was detected, and the density of Env protein and p24 bands was determined with an Odyssey Infrared Imaging system (LiCor Biosciences; Lincoln, NE). The level of Env protein in the transfected cells was expressed as fold differences over the amount of p24 Gag antigen in the same cells.

Sequence analysis

Sequences of SGA *env* amplicons were obtained by cycle-sequencing and dye terminator methods with an ABI 3730xl genetic analyzer (Applied Biosystems; Foster City, CA). Individual sequence fragments for each *env* SGA were assembled and edited using the Sequencer program 4.7 (Gene Codes; Ann Arbor, MI). The *env* sequences were aligned using CLUSTAL W (Thompson et al., 1994), and Genecutter (www.hiv.lanl.gov) for obtaining codon aligned files that could be used for signature analysis. Phylogenetic tree was constructed with complete *env* gene sequences using the neighbor-joining method (Saitou and Nei, 1987) and the Kimura two-parameter model (Kimura, 1980). The reliability of topologies was estimated by performing bootstrap analysis with 1000 replicates.

Hierarchical clustering analysis

The neutralizing potential of plasma (the log reciprocal dilution ID₅₀ scores) was organized into patterns of similar neutralizing potency when tested against the panel of envelopes used in this study using the Los Alamos National Lab web-based heatmap interface (<http://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html>) and the statistical package R (R Development Core Team, 2009). The test panel of Envs was resampled 10,000 times using a random-with-replacement bootstrap strategy, and this identified two robust and distinctive clusters recurring in 96% of the bootstrap samples: one group with little cross-neutralizing potential and another group with the ability to cross-neutralize multiple heterologous C clade viruses (Shimodaira, 2004). These two groups were used for subsequent signature analysis. The same strategy was used to identify statistically robust clusters of the Envs with like-neutralizing susceptibility. Clusters with high bootstrap values typically only included a few Envs each, and these Env clusters were typically found within a single subject. Therefore, we decided to organize Env pseudoviruses by individual in a heatmap figure (Fig. 7), while retaining the information regarding significant bootstrap clusters, shown superimposed onto the figure.

Signature analysis

We first divided the envelope sequences into the two groups, based on heatmap clustering patterns that indicated whether they were derived from an individual with plasma with a cross-reactive neutralizing profile or a weakly neutralizing profile. Alignments used for signature analysis were generated with GeneCutter (www.hiv.lanl.gov) to provide codon-aligned DNA for phylogenetic analysis. Phylogenetically corrected methods were used to identify all signature sites. Phylogenetic correction is critical because observed patterns in data can either result from correlations imposed by the initial historical emergence of a lineage of viruses (founder effects) or, in the case of HIV-1, be a consequence of recent biological interactions. Not accounting for founder effects can lead to erroneous statistical conclusions (Bhattacharya et al., 2007). The sequence of the virus depends on its full evolutionary history, while causal correlations are manifest in correlations with recent changes. The separation of the two effects, i.e., a phylogenetic correction, enables one to estimate the impact of recent changes on the phenotype. This requires statistical reconstruction of genealogical relationships between the viruses and an estimate of ancestral states of the viruses. We implemented this through maximum likelihood phylogenies and tested for phenotypic associations with mutational patterns based on change or stasis in a given amino acid, when compared to the most recent common ancestor, adapting the phylogenetically corrected Fisher's Exact Methods first developed by Bhattacharya et al. (2007). We screened the mutational pattern of each amino acid found at every column in the alignment. To correct for multiple tests, we used a q -value to assess the false discovery rate (Storey and Tibshirani, 2003). The four strongest associations had a p -value of 0.03 and q -value of 0.27, thus it is likely that at least one of the signatures identified is a false positive, and only borderline significance was obtained overall. Consequently, these signatures should be considered in a hypothesis-raising framework as potentially interesting sites that merit further investigation.

Nucleotide sequences accession numbers

GenBank accession numbers of the *env* gene sequences are GU329048–GU329523.

Supplementary material related to this article can be found online at [doi:10.1016/j.virol.2010.09.031](https://doi.org/10.1016/j.virol.2010.09.031).

Acknowledgments

This work was supported by grants from the National Institutes of Health/National Institute of Allergy and Infectious Diseases [CIPRA (R03 AI054155) to RMM, HIVRAD PO 1 AI35351, Center for HIV/AIDS Vaccine Immunology (AI067854), Duke Center for AIDS Research (AI64518) Molecular Virology Core], the Bill and Melinda Gates Foundation, and a Los Alamos National Laboratory directed research grant.

References

- Bhattacharya, T., Daniels, M., Heckerman, D., Foley, B., Frahm, N., Kadie, C., Carlson, J., Yusim, K., McMahon, B., Gaschen, B., Mallal, S., Mullins, J.L., Nickle, D.C., Herbeck, J., Rousseau, C., Learn, G.H., Miura, T., Brander, C., Walker, B., Korber, B., 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315, 1583–1586.
- Binley, J.M., Wrin, T., Korber, B., Zwick, M.B., Wang, M., Chappey, C., Stiegler, G., Kunert, R., Zolla-Pazner, S., Katinger, H., Petropoulos, C.J., Burton, D.R., 2004. Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J. Virol.* 78, 13232–13252.
- Binley, J.M., Lybarger, E.A., Crooks, E.T., Seaman, M.S., Gray, E., Davis, K.L., Decker, J.M., Wycuff, D., Harris, L., Hawkins, N., Wood, B., Nathe, C., Richman, D., Tomaras, G.D., Bibollet-Ruche, F., Robinson, J.E., Morris, L., Shaw, G.M., Montefiori, D.C., Mascola, J.R., 2008. Profiling the specificity of neutralizing antibodies in a large panel of plasmas from patients chronically infected with human immunodeficiency virus type 1 subtypes B and C. *J. Virol.* 82, 11651–11668.

- Bull, M.E., Learn, G.H., McElhone, S., Hitti, J., Lockhart, D., Holte, S., Dragavon, J., Coombs, R.W., Mullins, J.L., Frenkel, L.M., 2009. Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus. *J. Virol.* 83, 6020–6028.
- Burton, D.R., Desrosiers, R.C., Doms, R.W., Koff, W.C., Kwong, P.D., Moore, J.P., Nabel, G.J., Sodroski, J., Wilson, I.A., Wyatt, R.T., 2004. HIV vaccine design and the neutralizing antibody problem. *Nat. Immunol.* 5, 233–236.
- Cham, F., Zhang, P.F., Heyndrickx, L., Bouma, P., Zhong, P., Katinger, H., Robinson, J., van der Groen, G., Quinnan Jr., G.V., 2006. Neutralization and infectivity characteristics of envelope glycoproteins from human immunodeficiency virus type 1 infected donors whose sera exhibit broadly cross-reactive neutralizing activity. *Virology* 347, 36–51.
- Chertova, E., Bess Jr., J.W., Crise, B.J., Sowder, I.R., Schaden, T.M., Hilburn, J.M., Hoxie, J.A., Benveniste, R.E., Lifson, J.D., Henderson, L.E., Arthur, L.O., 2002. Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. *J. Virol.* 76, 5315–5325.
- Cordonnier, A., Montagnier, L., Emerman, M., 1989. Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. *Nature* 340, 571–574.
- Derdeyn, C.A., Decker, J.M., Bibollet-Ruche, F., Mokili, J.L., Muldoon, M., Denham, S.A., Heil, M.L., Kasolo, F., Musonda, R., Hahn, B.H., Shaw, G.M., Korber, B.T., Allen, S., Hunter, E., 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303, 2019–2022.
- Desrosiers, R.C., 2004. Prospects for an AIDS vaccine. *Nat. Med.* 10, 221–223.
- Dhillon, A.K., Donners, H., Pantophlet, R., Johnson, W.E., Decker, J.M., Shaw, G.M., Lee, F.H., Richman, D.D., Doms, R.W., Vanham, G., Burton, D.R., 2007. Dissecting the neutralizing antibody specificities of broadly neutralizing sera from human immunodeficiency virus type 1-infected donors. *J. Virol.* 81, 6548–6562.
- Diskin, R., Marcovaccio, P.M., Bjorkman, P.J., 2010. Structure of a clade C HIV-1 gp120 bound to CD4 and CD4-induced antibody reveals anti-CD4 polyreactivity. *Nat. Struct. Mol. Biol.* 17, 608–613.
- Fang, G., Zhu, G., Burger, H., Keithly, J.S., Weiser, B., 1998. Minimizing DNA recombination during long RT-PCR. *J. Virol. Meth.* 76, 139–148.
- Gao, F., Searce, R.M., Alam, S.M., Hora, B., Xia, S., Hohm, J.E., Parks, R.J., Ogburn, D.F., Tomaras, G.D., Park, E., Lomas, W.E., Maino, V.C., Fiscus, S.A., Cohen, M.S., Moody, M.A., Hahn, B.H., Korber, B.T., Liao, H.X., Haynes, B.F., 2009. Cross-reactive monoclonal antibodies to multiple HIV-1 subtype and SIVcpz envelope glycoproteins. *Virology* 394, 91–98.
- Gnanakaran, S., Lang, D., Daniels, M., Bhattacharya, T., Derdeyn, C.A., Korber, B., 2007. Clade-specific differences between human immunodeficiency virus type 1 clades B and C: diversity and correlations in C3–V4 regions of gp120. *J. Virol.* 81, 4886–4891.
- Gnanakaran, S., Daniels, M.G., Bhattacharya, T., Lapedes, A.S., Sethi, A., Li, M., Tang, H., Greene, K., Gao, H., Haynes, B.F., Cohen, M.S., Shaw, G.M., Seaman, M.S., Kumar, A., Gao, F., Montefiori, D.C., Korber, B., 2010. Genetic signatures in the envelope glycoproteins of HIV-1 that associate with broadly neutralizing antibodies. *PLoS Comput. Biol.* 6, e1000955.
- Haynes, B.F., Montefiori, D.C., 2006. Aiming to induce broadly reactive neutralizing antibody responses with HIV-1 vaccine candidates. *Expert Rev. Vaccin.* 5, 347–363.
- Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 20, W13–W23.
- Huang, C.C., Tang, M., Zhang, M.Y., Majeed, S., Montabana, E., Stanfield, R.L., Dimitrov, D.S., Korber, B., Sodroski, J., Wilson, I.A., Wyatt, R., Kwong, P.D., 2005. Structure of a V3-containing HIV-1 gp120 core. *Science* 310, 1025–1028.
- Huang, W., Toma, J., Fransen, S., Stawiski, E., Reeves, J.D., Whitcomb, J.M., Parkin, N., Petropoulos, C.J., 2008. Coreceptor tropism can be influenced by amino acid substitutions in the gp41 transmembrane subunit of human immunodeficiency virus type 1 envelope protein. *J. Virol.* 82, 5584–5593.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14 (33–38), 27–38.
- Irungi, J., Go, E.P., Zhang, Y., Dalpathado, D.S., Liao, H.X., Haynes, B.F., Desaire, H., 2008. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass Spectrom.* 19, 1209–1220.
- Kalia, V., Sarkar, S., Gupta, P., Montelaro, R.C., 2005. Antibody neutralization escape mediated by point mutations in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41. *J. Virol.* 79, 2097–2107.
- Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee, H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 105, 7552–7557.
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.
- Kirchherr, J.L., Lu, X., Kasongo, W., Chalwe, V., Mwananyanda, L., Musonda, R.M., Xia, S.M., Searce, R.M., Liao, H.X., Montefiori, D.C., Haynes, B.F., Gao, F., 2007. High throughput functional analysis of HIV-1 *env* genes without cloning. *J. Virol. Meth.* 143, 104–111.
- Korber, B., Gnanakaran, S., 2009. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. *Curr. Opin. HIV AIDS* 4, 408–417.
- Korber, B.T., Letvin, N.L., Haynes, B.F., 2009. T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces. *J. Virol.* 83, 8300–8314.

- Kulkarni, S.S., Lapedes, A., Tang, H., Gnanakaran, S., Daniels, M.G., Zhang, M., Bhattacharya, T., Li, M., Polonis, V.R., McCutchan, F.E., Morris, L., Ellenberger, D., Butera, S.T., Bollinger, R.C., Korber, B.T., Paranjape, R.S., Montefiori, D.C., 2009. Highly complex neutralization determinants on a monophyletic lineage of newly transmitted subtype C HIV-1 Env clones from India. *Virology* 385, 505–520.
- LaBranche, C.C., Sauter, M.M., Haggarty, B.S., Vance, P.J., Romano, J., Hart, T.K., Bugelski, P.J., Marsh, M., Hoxie, J.A., 1995. A single amino acid change in the cytoplasmic domain of the simian immunodeficiency virus transmembrane molecule increases envelope glycoprotein expression on infected cells. *J. Virol.* 69, 5217–5227.
- Li, M., Gao, F., Mascola, J.R., Stamatatos, L., Polonis, V.R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K.M., Bilski, M., Kothe, D.L., Salazar-Gonzalez, J.F., Wei, X., Decker, J.M., Hahn, B.H., Montefiori, D.C., 2005. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* 79, 10108–10125.
- Li, M., Salazar-Gonzalez, J.F., Derdeyn, C.A., Morris, L., Williamson, C., Robinson, J.E., Decker, J.M., Li, Y., Salazar, M.G., Polonis, V.R., Mlisana, K., Karim, S.A., Hong, C., Greene, K.M., Bilski, M., Zhou, J., Allen, S., Chomba, E., Mulenga, J., Vwalika, C., Gao, F., Zhang, M., Korber, B.T., Hunter, E., Hahn, B.H., Montefiori, D.C., 2006. Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J. Virol.* 80, 11776–11790.
- Li, Y., Migueles, S.A., Welcher, B., Svehla, K., Phogat, A., Louder, M.K., Wu, X., Shaw, G.M., Connors, M., Wyatt, R.T., Mascola, J.R., 2007. Broad HIV-1 neutralization mediated by CD4-binding site antibodies. *Nat. Med.* 13, 1032–1034.
- Lin, G., Nara, P.L., 2007. Designing immunogens to elicit broadly neutralizing antibodies to the HIV-1 envelope glycoprotein. *Curr. HIV Res.* 5, 514–541.
- Liu, S.L., Rodrigo, A.G., Shankarappa, R., Learn, G.H., Hsu, L., Davidov, O., Zhao, L.P., Mullins, J.L., 1996. HIV quasispaces and resampling. *Science* 273, 415–416.
- Moore, J.P., McCutchan, F.E., Poon, S.W., Mascola, J., Liu, J., Cao, Y., Ho, D.D., 1994. Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies. *J. Virol.* 68, 8350–8364.
- Moore, P.L., Gray, E.S., Choge, I.A., Rancobe, N., Mlisana, K., Abdool Karim, S.S., Williamson, C., Morris, L., 2008. The C3–v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *J. Virol.* 82, 1860–1869.
- Moore, P.L., Rancobe, N., Lambson, B.E., Gray, E.S., Cave, E., Abrahams, M.R., Bandawe, G., Mlisana, K., Abdool Karim, S.S., Williamson, C., Morris, L., 2009. Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog.* 5, e1000598.
- Morris, J.F., Sternberg, E.J., Gutshall, L., Petteway Jr., S.R., Ivanoff, L.A., 1994. Effect of a single amino acid substitution in the V3 domain of the human immunodeficiency virus type 1: generation of revertant viruses to overcome defects in infectivity in specific cell types. *J. Virol.* 68, 8380–8385.
- Palmer, S., Kearney, M., Maldarelli, F., Halvas, E.K., Bixby, C.J., Bazmi, H., Rock, D., Falloon, J., Davey Jr., R.T., Dewar, R.L., Metcalf, J.A., Hammer, S., Mellors, J.W., Coffin, J.M., 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 43, 406–413.
- R Development Core Team, 2009. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rademeyer, C., Moore, P.L., Taylor, N., Martin, D.P., Choge, I.A., Gray, E.S., Sheppard, H. W., Gray, C., Morris, L., Williamson, C., 2007. Genetic characteristics of HIV-1 subtype C envelopes inducing cross-neutralizing antibodies. *Virology* 368, 172–181.
- Rong, R., Gnanakaran, S., Decker, J.M., Bibollet-Ruche, F., Taylor, J., Sfakianos, J.N., Mokili, J.L., Muldoon, M., Mulenga, J., Allen, S., Hahn, B.H., Shaw, G.M., Blackwell, J.L., Korber, B.T., Hunter, E., Derdeyn, C.A., 2007. Unique mutational patterns in the envelope alpha 2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. *J. Virol.* 81, 5658–5668.
- Rong, R., Li, B., Lynch, R.M., Haaland, R.E., Murphy, M.K., Mulenga, J., Allen, S.A., Pinter, A., Shaw, G.M., Hunter, E., Robinson, J.E., Gnanakaran, S., Derdeyn, C.A., 2009. Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. *PLoS Pathog.* 5, e1000594.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Salazar-Gonzalez, J.F., Bailes, E., Pham, K.T., Salazar, M.G., Guffey, M.B., Keele, B.F., Derdeyn, C.A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J.A., Swanstrom, R., Haynes, B.F., Athreya, G.S., Korber, B.T., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* 82, 3952–3970.
- Sather, D.N., Armann, J., Ching, L.K., Mavrantoni, A., Sellhorn, G., Caldwell, Z., Yu, X., Wood, B., Self, S., Kalams, S., Stamatatos, L., 2009. Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *J. Virol.* 83, 757–769.
- Shen, X., Parks, R.J., Montefiori, D.C., Kirchherr, J.L., Keele, B.F., Decker, J.M., Blattner, W.A., Gao, F., Weinhold, K.J., Hicks, C.B., Greenberg, M.L., Hahn, B.H., Shaw, G.M., Haynes, B.F., Tomaras, G.D., 2009. In vivo gp41 antibodies targeting the 2F5 monoclonal antibody epitope mediate human immunodeficiency virus type 1 neutralization breadth. *J. Virol.* 83, 3617–3625.
- Shimizu, N., Haraguchi, Y., Takeuchi, Y., Soda, Y., Kanbe, K., Hoshino, H., 1999. Changes in and discrepancies between cell tropisms and coreceptor uses of human immunodeficiency virus type 1 induced by single point mutations at the V3 tip of the env protein. *Virology* 259, 324–333.
- Shimodaira, H., 2004. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Stat.* 32, 2616–2641.
- Shioda, T., Oka, S., Ida, S., Nukihara, K., Toriyoshi, H., Mori, S., Takebe, Y., Kimura, S., Shimada, K., Nagai, Y., et al., 1994. A naturally occurring single basic amino acid substitution in the V3 region of the human immunodeficiency virus type 1 env protein alters the cellular host range and antigenic structure of the virus. *J. Virol.* 68, 7689–7696.
- Stamatatos, L., Morris, L., Burton, D.R., Mascola, J.R., 2009. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat. Med.* 15, 866–870.
- Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100, 9440–9445.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 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.
- Trkola, A., Pomales, A.B., Yuan, H., Korber, B., Maddon, P.J., Allaway, G.P., Kattinger, H., Barbas III, C.F., Burton, D.R., Ho, D.D., et al., 1995. Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J. Virol.* 69, 6609–6617.
- Verhofstede, C., Demecheleer, E., De Cabooter, N., Gaillard, P., Mwanyumba, F., Claeys, P., Chohan, V., Mandaliya, K., Temmerman, M., Plum, J., 2003. Diversity of the human immunodeficiency virus type 1 (HIV-1) env sequence after vertical transmission in mother-child pairs infected with HIV-1 subtype A. *J. Virol.* 77, 3050–3057.
- Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrinn, T., Simek, M.D., Fling, S., Mitcham, J.L., Lehman, J.K., Priddy, F.H., Olsen, O.A., Frey, S.M., Hammond, P.W., Miir, G., Serwanga, J., Pozniak, A., McPhee, D., Manigart, O., Mwananyanda, L., Karita, E., Inwoley, A., Jaoko, W., Dehovitz, J., Bekker, L.G., Pitisuttithum, P., Paris, R., Allen, S., Kaminsky, S., Zamb, T., Moyle, M., Koff, W.C., Poignard, P., Burton, D.R., 2009. Broad and potent neutralizing antibodies from an African donor reveal a New HIV-1 vaccine target. *Science* 326, 285–289.
- Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D., Shaw, G.M., 2003. Antibody neutralization and escape by HIV-1. *Nature* 422, 307–312.
- Wu, X., Yang, Z.-Y., Li, Y., Hogerkor, C.-M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., Longo, N.S., McKee, K., O'Dell, S., Louder, M.K., Wycuff, D.L., Feng, Y., Nason, M., Doria-Rose, N., Connors, M., Kwong, P.D., Roederer, M., Wyatt, R.T., Nabel, G.J., Mascola, J.R., 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329, 856–861.
- Zhang, P.F., Cham, F., Dong, M., Choudhary, A., Bouma, P., Zhang, Z., Shao, Y., Feng, Y.R., Wang, L., Mathy, N., Voss, G., Broder, C.C., Quinlan Jr., G.V., 2007. Extensively cross-reactive anti-HIV-1 neutralizing antibodies induced by gp140 immunization. *Proc. Natl. Acad. Sci. USA* 104, 10193–10198.
- Zhou, T., Georgiev, I., Wu, X., Yang, Z.Y., Dai, K., Finzi, A., Do Kwon, Y., Scheid, J., Shi, W., Xu, L., Yang, Y., Zhu, J., Nussenzweig, M.C., Sodroski, J., Shapiro, L., Nabel, G.J., Mascola, J.R., Kwong, P.D., 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* 329, 811–817.
- Zhu, P., Chertova, E., Bess Jr., J., Lifson, J.D., Arthur, L.O., Liu, J., Taylor, K.A., Roux, K.H., 2003. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. USA* 100, 15812–15817.
- Zwick, M.B., Labrijn, A.F., Wang, M., Spenlehauer, C., Saphire, E.O., Binley, J.M., Moore, J.P., Stiegler, G., Kattinger, H., Burton, D.R., Parren, P.W., 2001. Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *J. Virol.* 75, 10892–10905.