Selective regimes and evolutionary rates of HIV-1 subtype B V3 variants in the Brazilian epidemic

Ricardo Sobhie Diaz a,⁎, Élcio Leal a, Sabri Sanabani a,b, Maria Cecilia A. Sucupira a, Amílcar Tanuri c, Ester C. Sabino b, Luiz Mário Janini a

a Federal University of São Paulo, Brazil
b Fundação Pró-Sangue / Hemocentro de São Paulo, Brazil
c Federal University of Rio de Janeiro, Brazil

A R T I C L E   I N F O

Keywords:
HIV-1 selective regimes
Genetic diversity
V3 envelop region
Clade B
Brazil

A B S T R A C T

Half of subtype B Brazilian HIV-1 harbors the V3 tip GWGR instead of the GPGR. To investigate the evolution of GW variants, we analyzed 81 env sequences and 5 full-length GW genomes from antiretroviral-naïve individuals sampled between 1983 and 1999. Phylogenetic analysis indicated that GW strains intermingle in the tree with other subtype B sequences. The mean \( d_N/d_S \) values of GW strains were proximal to those of the other sequences, regardless of sampling years or clinical status. In sequences from patients with CD4+ T cell counts ≥200 cells/μL, the mean \( d_N/d_S \) ratio was greater than one, suggesting a positive selection. The prevalence of GW variants was lower among individuals in whom disease progressed. This is probably attributable to the fact that tryptophan is replaced by other amino acids over time, whereas the GP motif does not evolve as rapidly.

© 2008 Elsevier Inc. All rights reserved.

Introduction

The potential of HIV-1 to generate genetic diversity is remarkable. This diversity results from the elevated rate at which the viral reverse transcriptase incorporates errors during the numerous replication events occurring in infected patients (Ho et al., 1995; Roberts, Bebenek, and Kunkel, 1988; Wei et al., 1995). At the global level, HIV-1 strains are classified as belonging to one of three groups (M, N or O), nine subtypes (A, B, C, D, F, G, H, J or K) and 34 circulating recombinant forms (http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html). In Brazil, HIV-1 subtypes B, F and C, as well as unique circulating recombinant forms, are co-circulating (Bongertz et al., 2000; Csillag, 1994; De Sa Filho et al., 2006; Louwagie et al., 1994; Morgado et al., 1998; Morgado et al., 1994; Potts et al., 1993; Ramos et al., 1999; Sa Filho et al., 2005; Santos et al., 2006; Thomson et al., 2004). In 1993 (Potts et al., 1993), it became apparent that Brazilian subtype B strains contained an antigenically distinct variant that has a signature motif at the tip of the V3 loop, with tryptophan rather than proline at position 328 in the HIV-1 envelope (Casseb et al., 1998; da Costa et al., 1995; Potts et al., 1993). It has consistently been shown that this variant accounts for approximately half of the subtype B strains circulating in Brazil and is exceedingly rare in other parts of the world (Bongertz et al., 2000; Casseb et al., 1998; Morgado et al., 1994). Analysis of Brazilian GWGR C2-V3-C3 and gp120 env sequences indicated they are closely related to North American and European subtype B strains, presenting 89.1% nucleic acid similarity to the North American MN strain (da Costa et al., 1995), although a tryptophan-to-proline substitution at this position requires a three-nucleotide substitution. Interestingly, in two separate studies, conducted in 2000 and 2002, respectively, it was suggested that GWGR strains possess particular biological properties, causing the pace of disease progression to be slower in individuals infected with GWGR strains than in those infected with the more common Brazilian GPGR strains (Casseb et al., 2002; Santoro-Lopes et al., 2000).

In order to further investigate GWCR strains in Brazil, we analyzed sequences of HIV subtype B presenting various motifs at the tip of the V3 loop. The data related to these different sequences were divided into four sets, by the group to which they relate: BR-83 (the earliest available sequences in the Brazilian HIV epidemic, 15 sequences obtained in 1983); BR-94 (25 sequences obtained from asymptomatic patients sampled in 1994); BR-95 (26 sequences obtained from symptomatic patients sampled in 1995); and BR-99 (15 sequences obtained from untreated asymptomatic patients collected in 1999). All samples were obtained from antiretroviral-naïve individuals and therefore represent the natural viral evolution of HIV in relation to the progression of the disease. In addition, five full-length genomes from recently infected individuals harboring the GWGR motif at the tip of the V3 loop...
were compared to other, previously described, full-length genomes of clade B strains.

Results

The HIV-1 prevalence in the MSM population evaluated in 1983 was 36% (24 of the 66 patients tested). Using the dual EIA testing strategy, we found that 44% of the individuals in the 1983 sample presented recent HIV-1 infection (3A11-positive/3A11-DT-negative). The annual incidence in this population was 59.8%. With the exception of the 5 HIV-1 negative controls, all of the samples were 3A11-positive. The ODs were equivalent in all 6 duplicate pairs of samples. None of the 4 follow-up samples collected in 1986 were 3A11-DT-negative. Sample BZ-328 (collected in 1983) was 3A11-DT-negative, and the follow-up sample collected in 1986 from the same patient (sample BZ-380) was 3A11-DT-positive, showing the progressive increase in antibody titers over time in this patient.

We were able to amplify and sequence HIV from the sera of 15 of the 24 1993 samples. Of the 24 serum samples collected in 1983, 12 (50%) presented GW-specific peptide binding, 10 (42%) presented GP-specific peptide binding, and 8% presented an indeterminate status. It is of note that, in one sample presenting GPGR in the sequence analysis, GWGR peptide-binding antibodies were identified. With this one exception, there was congruity between the sequences obtained and the peptide-binding specificity. Sequences were also obtained from the 3 follow-up samples, thereby confirming the relationship between sequences obtained in 1983 and 1985 (data not shown). Two sequences in which the rare tetramers GAGR and GVGR were detected were later classified as GWGR in the GP-specific peptide-binding assay and, in the STARHS, were not categorized as representing recent infections. Excluding those 2 samples, the sequencing results showed that 6 individuals harbored the GPG motif and 7 harbored the GWG motif. Interestingly, 8 samples collected in 1994 or 1995 and harboring other amino acids at the P/W position, such as A (gcg), V (gtg), E (gag), F (ttt), and L (ttg), all reacted as W in the peptide-binding assay. As can be seen in Table 1, the prevalences of P and W vary according to the individual CD4+ T cell counts, the prevalence of W being higher among individuals with less evidence of progression (CD4 counts <200 cells/μL; chi-square p <0.03). However, if we considered that amino acids other than P or W are evolving from W, there would be no statistical differences between the two groups (see further explanation below).

In order to investigate the possible evolutionary relationships among GWGR samples, a V3 env maximum likelihood tree was constructed using all of the sequences in the four data sets (BR-83, BR-94, BR-95 and BR-99). In this tree (data not shown), GWGR sequences tended to cluster, sequences with other V3 motifs grouping far from the GWGR sequences. Since the fragments generated had only 216 bp,

![Fig. 1. HIV-1 C2-C3 env maximum likelihood tree constructed from a 204-bp fragment.](image-url)

The Brazilian sequences are represented by circles, whereas HIV-I group M reference sequences are represented by diamonds. The blue circles represent samples obtained from asymptomatic patients, red circles represent samples obtained from symptomatic patients with CD4 counts <200 cells/μL, and magenta circles represent samples obtained from symptomatic patients with CD4 counts ≥200 cells/μL. The V3 motifs of the samples are represented to the right of the tree (only motifs other than GPGR are shown). (For interpretation of the figure legend, the reader is referred to the web version of this article.)

Table 1

<table>
<thead>
<tr>
<th>CD4+ T cell cells/μL</th>
<th>Amino acid at the V3 tip tetramer</th>
<th>Proline</th>
<th>Tryptophan</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥200</td>
<td></td>
<td>17 (52%)</td>
<td>8 (24%)</td>
<td>8 (24%)</td>
<td>33</td>
</tr>
<tr>
<td>≤200</td>
<td></td>
<td>11 (33%)</td>
<td>18 (54%)</td>
<td>4 (12%)</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>28 (43%)</td>
<td>26 (39%)</td>
<td>12 (18%)</td>
<td>66</td>
</tr>
</tbody>
</table>
it was unclear whether the grouping tendency of GWGR samples was influenced by the 12-bp V3 tip motif. Therefore, a second maximum likelihood tree was constructed using the same sequences without the V3 tip tetramer. In this second tree (Fig. 1), sequences bearing GWGR, GPGR or other V3 motifs did not establish clusters; rather, they were all evenly dispersed throughout the tree. Therefore, other than the V3 tip signature, there were no mutations supporting the clustering of GWGR samples in this fragment. In addition, we did not observe any grouping in Brazilian subtype B V3 sequences, regardless of whether they were sampled earlier (1983) or later (1999) in the epidemic or whether they derived from symptomatic or asymptomatic patients. However, in this second tree, the branches related to samples collected in 1995 from symptomatic patients with CD4 counts > 200 cells/μL (Fig. 1) were longer than those related to other samples within the same phylogenetic pair.

Owing to the limited (216-bp) env fragment of the previous tree, we decided to perform further analysis using a 746-bp env fragment from Brazilian samples obtained in 1999, by this time maintaining the V3 tip motif. Although the sequences with distinct motifs were included in this analysis, they did not form a cluster sharing the same V3 loop tetramer. Specifically, GWGR variants were evenly dispersed throughout this tree. Nevertheless, Brazilian sequences had a tendency to group together, away from the subtype B reference strains obtained from other regions of the world (Fig. 2).

To further characterize the relationship between the GW variants and other HIV-1 subtype B strains, we analyzed the full-length genomes of six variants generated in this study in an attempt to identify any GW related genetic signatures. Five of these sequences presented GWGR at the V3 tip. For comparative purposes, we also included in this analysis 50 full-length sequences of subtype B isolates from other countries (obtained from the Los Alamos HIV database). Those sequences included one full-length genome of a GW virus previously isolated in Brazil (O2BR008 GenBank ID: DQ358805) and one of a GW virus previously isolated in Argentina (04AR143170 GenBank ID: DQ383750). In the tree, the sequences isolated in Brazil (with the exception of sequence B.BR.89.BZ167, isolated in 1989) were included in a well-defined cluster with high a posteriori probability (Fig. 3). This cluster included sequences from Brazil, Argentina and Uruguay, suggesting that there is significant cross-infection among these countries. Notably, despite the fact that all isolates presenting the GWGR tetramer, as well as the majority of all Brazilian isolates, are monophyletic, they did not form a unique cluster. This suggests that the current HIV-1 subtype B variants in Brazil originated from the same ancestral sequence.

In order to perform a more in-depth analysis of Brazilian V3 env sequences containing different signatures, we calculated mean pairwise distances. To that end, we separated sample data into data sets by sampling year (BR-83, BR-94, BR-95 and BR-99, as previously mentioned), by V3 tetramer signature (BR-GPGR, BR-GWGR and others) and by patient clinical status (symptomatic and asymptomatic). Mean pairwise distances related to the V3 loop tetramer ranged from 0.14 to 0.24. Similarly, mean pairwise distances related to the sampling year ranged from 0.13 to 0.20, sequences isolated in 1983 (BR-83) presenting the least intragroup distance and those isolated in 1995 (BR-95) presenting the greatest (Table 2). Mean pairwise distances related to patient clinical status ranged from 0.16 to 0.21, data sets from symptomatic individuals presenting the highest values. Data related to symptomatic patients were then further divided into two data subsets: Sympt-SP, including sequences from patients with CD4 counts ≤ 200 cells/μL, and Sympt-AP, including sequences from patients with CD4 counts of 200 to 300 cells/μL. Interestingly, even within this symptomatic category, we noticed that the mean pairwise distance was greater in the data subset Sympt-AP (Table 2). Although the previous analyses revealed subtle differences among the data sets, these results should be evaluated with caution, since the pairwise method is highly sensitive to the number and length of sequences.

In addition, pairwise estimates of \( d_{NS}/d_{S} \) ratios were also performed in order to determine the selective regimes of the V3 loop variants. Accordingly, \( d_{NS}/d_{S} \) ratios (\( \omega \) values) were not significantly different than 1 for all data, and we therefore could not reject the hypothesis of neutral evolution (Table 2). Despite the lack of statistical support (probably due to the small sample size and the shortness of the
fragments analyzed), the observed inflation of the $\omega$ value in symptomatic individuals with high CD4 counts merited further study. Therefore, we also analyzed the symptomatic individuals according to their CD4 counts using a codon-based maximum likelihood method. The results of this analysis are summarized in Table 3; the likelihood ratio test indicated striking differences regarding the selective regime. Specifically, in patients with CD4 counts $\leq$200 cells/µL, the number of conserved sites ($\omega=0$) was similar to that of neutral sites ($\omega=1$) according to the neutral model (model 1). However, according to the discrete model (model 3), in patients with CD4 counts ≥200 cells/µL, 23% of the sites were conserved, 40% were under purifying selection ($\omega=0.35099$), and 37% were under positive selection ($\omega=1.61380$). In addition, model 3 identified eight codons (33, 36, 45, 59, 62, 64, 65 and 67) as having the highest a posteriori probability (0.99) of positive selection in the 216-bp C2-V3-C3 fragment (Table 3 and Fig. 4). Most of these codons were located either in the vicinity of or within the V3 tip, codon 36 (the second codon of the tetramer) being the only codon located within the tip (Fig. 4).

Discussion

The HIV-1 subtype B strains continue to be the dominant viruses in the Brazilian epidemic (Sa Filho et al., 2005). However, some variants harbor the GWGR tetramer, which is quite antigenically distinct, at the tip of the V3 loop (Casbe et al., 1998; Potts et al., 1993). These variants account for approximately half of all subtype B-related infections in Brazil (Casbe et al., 2002; Potts et al., 1993; Santoro-Lopes et al., 2000). It has been speculated that these variants are associated with slower disease progression (Casbe et al., 2002; Morgado et al., 1994; Potts et al., 1993; Santoro-Lopes et al., 2000). Notably, we confirmed that the GWGR variants were present at a very early stage of the Brazilian epidemic, as seen in the 1983 samples. Together with other facts, the finding of GWGR sequences at this early stage suggests a founder effect, rather than a constant evolution from early stage of the Brazilian epidemic, at which samples were obtained. Although we analyzed samples collected between 1983 and 1999 (spanning 16 years of the epidemic), no major differences in branch length or phylogenetic segregation were observed. It should be noted that both analyses were based on env fragments, with limited sequence information due to the short length of the fragments analyzed. Therefore, we performed additional analyses of five full-length genome sequences of GW viruses isolated in Brazil. Similarly, we found that GW viruses did not form a unique cluster and were not distinguishable from other full-length genome sequences of clade B viruses previously isolated in Brazil. The full-length tree also indicated that, with the exception of one sample (BR.89.BZ167), all of the sequences of GW viruses isolated in Brazil might have a common ancestor, regardless of the signature in the tetramer of V3 loop region of env gene.

Brazilian env sequences were further investigated by measuring mean pairwise distances and calculating $d_{\omega}/d_{L}$ ratios. In that analysis, sequences were grouped into different data sets by V3 tip motif (GPGR, GWGR and others) and by sampling date (1983, 1993, 1994 and 1999). Sequences were also grouped by patient clinical status at the moment samples were collected (symptomatic or asymptomatic). Mean pairwise distances related to V3 tip motif ranged from 0.14 (in the BR-GWGR data set) to 0.24 (in the data set consisting of sequences presenting neither GWGR nor GPGR motifs). The greater mean distance observed in the latter data set might be due to the presence of distinctive V3 signatures (APGR, GRGR, GGGR, etc), which would have also caused an increase in the standard error, or to a longer period of evolution of these strains. The mean pairwise distance in the BR-GPGR data set was 0.16, compared with 0.13 in the BR-GWGR data set. Mean pairwise distances in the data sets grouped by sampling date did not vary greatly, ranging from 0.13 in the BR-83 data set to 0.20 in the BR-95 data set. These results demonstrate that, although the variation is not great, there is a trend toward lower genetic diversity early in the epidemic and in the GWGR group of sequences. In addition, the genetic distances were greater among symptomatic patients than among asymptomatic patients. Mean pairwise distance obtained for the symptomatic patient data set was 0.16, whereas it was 0.19 and 0.21, respectively, in the Sympt-SP ($\leq$200 cells/µL) and Sympt-AP ($\geq$200 cells/µL) data subsets.

Pairwise $d_{\omega}/d_{L}$ ($\omega$) estimates indicated that all the abovementioned data sets were under selective regimes close to neutrality, since the test indicated $\omega$ values not different from 1 (Table 2). Pairwise distances and $d_{\omega}/d_{L}$ estimates suggested that Brazilian subtype B sequences do not have any V3 variants under specific selective regimes, nor do they present distinct evolutionary rates. Although based on data related to short sequence fragments, our study did not identify any phylogenetic or genetic features distinguishing GWGR variants from other subtype B variants circulating in Brazil. Therefore, observations made in other studies implying that GWGR variants harbor particular biological features that result in slower disease progression might not be unrelated to the genetic signatures and differential evolutionary rates of the env gene. Interestingly, our pairwise analysis results indicate that the $\omega$ value was slightly higher (1.23) in the patients with CD4 counts $\geq$200 cells/µL. Although not statistically different from 1 according to the t-test, this slightly higher $\omega$ value reveals a tendency toward $d_{\omega}$ elevation (Table 2). Since these sequences contained distinct V3 variants (Fig. 1), the higher $\omega$ values might reflect patient immune status rather than the presence of a specific V3 signature. These results, together with the finding that branch lengths were longer in sequences obtained from symptomatic patients with CD4 counts $\geq$200 cells/µL than in other sequences within the same phylogenetic pairs (Fig. 1), led us to perform a more detailed analysis of the selective regimes in these sequences.

We used a codon-based maximum likelihood method to compare the selective pressures among symptomatic patients separated by CD4 counts into the subsets of those with CD4 counts $\leq$200 cells/µL and those with CD4 counts $\geq$200 cells/µL. The results of this analysis
showed striking differences between these two data subsets in terms of the selective regimes. Specifically, for those with CD4 counts <200 cells/μL, the best-fit model (model 1) indicated that 56.5% of codons were conserved (ω = 0) and 43.5% were under neutral regime (ω = 1). However, for the patients with CD4 counts ≥200 cells/μL, the best-fit model (model 3) indicated that 37% of codons under positive selection (ω = 1.6). In addition, the latter model identified eight codons with a high (99%) a posteriori probability of being under positive selection (Fig. 3). Although those codons were located near the V3 loop, only the codon at the second position of this tetramer (codon 63; Fig. 3) was positively selected. It is noteworthy that, within the data subset related to patients with CD4 counts ≥200 cells/μL, various V3 motifs were identified. It is quite reasonable to suggest that positive selection is the main driving force of V3 loop tetramer diversification. In summary, the relatively greater genetic diversity in the group of symptomatic patients with higher CD4 counts, taken together with the higher dN/dS ratios, suggests that there is greater selective pressure exerted by a more preserved immune system even in the presence of AIDS-defining conditions.

Our findings indicate that Brazilian subtype B strains presenting GWGR tetratrams at the tip of the V3 loop were neither phylogenetically nor genetically distinct from those presenting GPGR tetramers. Codon-based maximum likelihood analyses identified codons under positive selection in the vicinity of and within the V3 tetramer in patients with CD4 counts ≥200 cells/μL, thereby showing that, even during the symptomatic period, the immune system readily recognizes and intensively responds to viral proteins. Finally, the fact that the frequency of GWGR variants remains high in Brazil is likely due to complex combinations of stochastic and deterministic forces such as founder effects, host cell factors and immune pressure. Since the prevalence of GWGR strains is the same as that of GPGR strains among individuals with higher CD4 counts, the transmission fitness of this variant is probably high. The great majority of HIV transmissions will occur close to the primary/recent infection (Brenner et al., 2007), and infecting individuals will therefore continue to harbor significant loads of the GWGR strain during this period. As disease progresses, the probability for tryptophan (Trp) disappearance increases, being replaced to closer amino acids such as G (Ggr), A (gcr), V (gtr), E (ggr), R (agr) and L (rgt), which are frequently found in Brazilian individuals with clade B strains and low CD4+ T cell counts. Using serological methods, we found that individuals harboring the abovementioned GWGR-related sequences will react as GWGR-infected individuals in the peptide-binding assay, suggesting that the GWGR sequence has been present in these individuals in the past. Other evidence, especially the fact that, in individuals with lower CD4+ T cell counts, the prevalence of GWGR sequences was found to decrease as the prevalence of sequences harboring other motifs increases (Table 1), corroborates this hypothesis. This, together with our finding that the mean pairwise distance is less in the GWGR sequences than in the GRGR sequences, suggests that genetic evolution is relatively more rapid in the GWGR sequences. Other interesting insights into the evolution from GWGR to other motifs are found in a study describing a patient failing antiretroviral therapy and harboring HIV strains with GGR motif at the tip of the V3 loop, X4 related strains, and antiretroviral resistant strains (Silva et al., 2006). That patient presented the de novo appearance of GWGR, R5, and wild type viruses after antiretroviral interruption. This reversion to ancient viruses driven by the interruption of antiretroviral therapy, as seen in other study (Wang et al., 2007), further suggests that sequences presenting GGR at the tip of the V3 loop evolve from GWGR sequences.

We recognize the fact that the small number of sequences analyzed here and the lack of a prospective cohort of infected individuals precludes the drawing of definite conclusions. Nevertheless, we believe that speculative studies of disease progression in GPGR-infected individuals in comparison with GWGR-infected individuals should be performed using the genetic characterization of the viruses and should take into account the possibility that tryptophan substitution occurs during the course of the infection.

**Material and methods**

**Population studied**

In a study of sexual transmission of hepatitis B conducted in 1983 at the Federal University of São Paulo School of Medicine, in São Paulo, Brazil, samples of sera were obtained from 66 men who have sex with men (MSM). These samples were aliquoted and stored undisturbed until gathered for retrospective analysis in the current study. Of the 66 samples analyzed, 24 (36%) were found to be anti-HIV-positive by enzyme immunoassay (EIA) and Western blot. All HIV-seropositive patients were asymptomatic for HIV infection in 1983. Follow-up samples collected in 1986 from 4 of those same individuals were also analyzed as controls for amplification carryover. Fifteen of the 66 samples yielded a quantity of HIV RNA sufficient for sequencing. Follow-up samples collected in 1985 from 3 other individuals in the same group were also analyzed and served as quality controls for the
samples collected in 1983. We also analyzed three additional sets of samples. The second set consisted of 25 serum samples obtained in 1994 from asymptomatic blood donors. The third set was composed of DNA samples collected in 1999 from 15 asymptomatic, untreated HIV-1-positive patients. An additional 5 samples collected in 2004 from recently infected individuals presenting the GWGR motif at the tip of the V3 loop were used for the full-length genome analysis. All samples presented here are from individuals not receiving antiretroviral therapy and therefore represent the natural evolution of HIV within the hosts.

Determining HIV-1 incidence

Using the serologic testing algorithm for recent HIV seroconversion (STARHS) ([Janssen et al., 1998]), we determined the incidence rate in the 1983 sample set. As this is the oldest set of samples of HIV-1 infected individuals in Brazil, we understood it would be interesting to evaluate the HIV incidence at that time in a country considered to be the epicenter of HIV epidemics in South America. The STARHS consists of a serologic testing algorithm to distinguish persons with early HIV-1 infection from other HIV-infected persons. The algorithm is based on the facts that (i) antibodies to HIV increase progressively over several months in recently infected individuals, and (ii) this period can be lengthened by modifying an EIA to make it less sensitive. For the more sensitive assay, we used a whole viral lysate-based EIA for HIV-1 (3A11, Abbot Laboratories, Chicago, IL, USA) according to the manufacturer instructions. For the less sensitive EIA, the following elements of the 3A11 test were modified as previously described ([Janssen et al., 1998]): sample dilution, sample incubation time, and conjugate incubation time. To further reduce the sensitivity in the less sensitive version of the 3A11, the cutoff for positivity in the less sensitive version of the 3A11, the cutoff for optical density (OD) was also increased to 0.7. This strategy was validated by testing 690 serial samples from 104 seroconverting individuals using the sensitive EIA (3A11) and the less sensitive version of the same assay (3A11-DT). It was determined that the mean interval between the point at which the 3A11 (reactive) assay can identify an individual with early HIV infection through evaluation of a blood specimen and that at which the 3A11-DT (nonreactive) assay can thus identify such an individual is 129 days (95% confidence interval=109–149 days). Annual incidence was estimated as follows:

\[ I = \frac{(ndt/N) \times (365/w)}{\text{where } I \text{ is the incidence (seroconversions per person–year), } ndt \text{ is the number of persons with 3A11-positive/3A11-DT-negative results, } N \text{ is the number of susceptible HIV-negative persons, and } w \text{ is the estimated mean time between seroconversion determined using the 3A11 assay and that determined using the 3A11-DT assay (129 days).}} \]

We analyzed, in a blinded fashion, a panel of 39 samples that included the 24 available samples from the 25 collected from HIV-1-infected MSM in 1983, 6 duplicate samples from those 24, 4 follow-up samples collected from those same individuals in 1986, and 5 HIV-negative control samples drawn in 1983.

### Table 3

Maximum likelihood analysis of $d_0$ and $d_3$ in the 216 bp HIV-1 C2–C3 envelope fragment

<table>
<thead>
<tr>
<th>Codon Model</th>
<th>-lnL</th>
<th>LRT</th>
<th>Positive selected sites with $p_2 = 0.99$</th>
<th>Parameters estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset: Sympt-SP ($n=13 \rightarrow 216$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>1329.395841</td>
<td>M0 vs M2 = 90.57, $p = 0.0000$</td>
<td>Not allowed</td>
<td>$k = 3.02018 \times 10^{-2}$</td>
</tr>
<tr>
<td>M2</td>
<td>1827.986512</td>
<td>M1 vs M2 = 7.75, $p = 0.0020$</td>
<td>Not allowed</td>
<td>$e_2 = 0.00000 \times 10^{10000}$</td>
</tr>
<tr>
<td>M3</td>
<td>1428.104149</td>
<td>M2 vs M3 = 1.54, $p = 0.0019$</td>
<td>59</td>
<td>$p = 0.53287 \times 10^{0.41}$</td>
</tr>
<tr>
<td>M4</td>
<td>1283.338243</td>
<td>M2 vs M3 = 1.54, $p = 0.0019$</td>
<td>59</td>
<td>$p = 0.53287 \times 10^{0.41}$</td>
</tr>
<tr>
<td>M7</td>
<td>1287.267990</td>
<td>M7 vs M8 = 7.39, $p = 0.0024$</td>
<td>Not allowed</td>
<td>$e_2 = 0.00000 \times 10^{10000}$</td>
</tr>
<tr>
<td>M8</td>
<td>1283.568421</td>
<td>M7 vs M8 = 7.39, $p = 0.0024$</td>
<td>59</td>
<td>$p = 0.98051 \times 0.45320 \times 0.61815$</td>
</tr>
</tbody>
</table>

* Best-fit model according to the LRT. $p$ = proportion of codons under a defined category of $w$ ($d_0$/$d_3$) value. Sympt-SP=symptomatic patients with CD4 levels below 200 cells/μL. Sympt-AP=symptomatic patients with CD4 levels above 200 cells/μL.
Peptide-binding assay

As shown in Table 4, synthetic biotinylated peptides based on the V3 region of HIV gp120 were used. Microplates (Immulon II round bottom; Dynatech Laboratories) were coated with avidin DX diluted in PBS, pH 7.4, at 10 μg/mL overnight at 4 °C. The peptides were biotinylated and were diluted 1 μg/mL in PBS, pH 7.4, 0.1 mL/well and incubated overnight at 4 °C. The plates were washed five times with PBS/0.05% Tween-20 and blocked with PBS plus 5% dry milk without Tween for 1 h at 37 °C, after which they were washed five additional times with PBS/Tween. Serum dilutions in PBS plus 5% dry milk plus 0.05% Tween-20 (Blotto) were then added (0.1 mL/well, starting at 1:50), and this was followed by a 1-h incubation at 37 °C. All runs included a human HIV-1-positive plasma pool as control. Each plate was run in duplicate: one replicate plate was washed five times with PBS-Tween 20; the other was washed five times with 8 M urea in PBS. Peroxidase labeled anti-human IgG (diluted 1:10,000 in Blotto) was added (0.1 mL/well), and the plates were incubated for 1 h at 37 °C, after which they were washed 5 times with PBS/Tween. The plates were then incubated with 0.15 mL/well of TMB substrate for 30 min at room temperature. The reaction was stopped with 0.50 mL/well of 4 N H2SO4 solution. The plates were read at 450 nm, and the endpoint titer was calculated as the dilution producing a mean absorbance of 0.5.

The scheme for classifying the motifs as W or P was as follows. Four ratios of the arithmetic titer ratios were calculated: BR1/SF2; BR1/MN; BR2/SF2 and BR2/MN'. The sum of all four ratios was computed. Sera presenting a sum ratio <8 were classified as P, whereas those presenting a sum ratio between 8 and 16 were classified as indeterminate, and those presenting a sum ratio >16 were classified as W.

Viral RNA isolation and reverse transcription PCR

To extract RNA, the Boom et al. method was applied, as previously described (Boom et al., 1990), to ≤200 μL of serum. The RNA was then converted into cDNA by reverse transcription (RT)-PCR (Gibson, Mori, and Clewley, 1993). In brief, RNA was recovered by binding to silica in the presence of guanidinium thiocyanate and subsequent elution in 40 μL of RNase-free water. Twenty microliters of the eluted RNA were combined with 10 pmol of an HIV-specific antisense primer (Wolfs et al., 1990), denatured for 10 min at 65 °C and immediately placed on ice. The mixture was incubated at 42 °C for 45 to 60 min in 40 μL of a solution containing 50 mM Tris–HCl, 40 mM KCl, 5 mM MgCl2, 0.5% Tween 20, 10 mM dithiothreitol, 20 U of RNasin (catalog no. N251a; Promega), 50 U of Expand Reverse Transcriptase (catalog no. 1785834; Boehringer Mannheim) and each deoxynucleoside triphosphate at a concentration of 1 mM. The reaction was stopped by putting the mixture on ice, after which the cDNA for the C2–V3–C3 region (designated the V3 region for simplicity) of the env gene was immediately amplified using nested PCR. The nested PCR was performed in 10 mM Tris (pH 8.3), 50 mM KCl, 3 mM MgCl2, 200 μM of each deoxynucleoside triphosphate, 0.2 μM of each primer and 2.5 U Taq polymerase (Perkin Elmer) per 100 μL of reaction mix. In the second round of nested PCR, 5 μL of the first-round PCR product was used. Nested amplification of the cDNA for the V3 region was performed using primers described elsewhere (Wolfs et al., 1990).

Table 4
Biocinated peptides used in the peptide-binding assay

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Peptide</th>
<th>Homology (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR1</td>
<td>NTRKSIH</td>
<td>100</td>
</tr>
<tr>
<td>BR2</td>
<td>M</td>
<td>93</td>
</tr>
<tr>
<td>SF2</td>
<td>Y</td>
<td>80</td>
</tr>
<tr>
<td>MN</td>
<td>P</td>
<td>80</td>
</tr>
<tr>
<td>3B</td>
<td>R</td>
<td>60</td>
</tr>
</tbody>
</table>

BR1 and BR2 are the consensus sequences for the Brazilian strain containing the GWGR motif at the tip of the V3 loop. SF2, MN, and 3B are American strains, and all strains used are from HIV-1 clade B. Points indicate amino acid identities, whereas dashes indicate gaps.
Thermal cycling conditions for V3 were 35 cycles of 95 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min, with a final extension at 72 °C for 10 min. Full-length genomes of the samples collected in 2004 from the 5 recently infected individuals were amplified and sequenced as previously described (Santos et al., 2006).

**Sequencing**

Purified PCR products were sequenced bidirectionally using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase (Applied Biosystems). The reaction products were sequenced and analyzed using an automated sequencer (ABI 3100; Applied Biosystems). Sequences were edited using the Sequencher 4.0 software (GeneCodes). The GenBank accession numbers of sequences are FJ195020 to FJ195091.

**Alignment and phylogenetic inference**

The sequences were manually aligned using the Se-Al program (version 2.0; Department of Zoology, Oxford University [http://evolve.zoo.ox.ac.uk/software/]). Sequences with insertions/deletions and stop codons were excluded from the alignments. Sets of HIV-1 subtype reference sequences obtained from the GenBank were included in the alignments. Phylogenetic inferences were made using the general time-reversible nucleotide substitution model (Rodriguez et al., 1990). Maximum likelihood trees were first determined by the neighbor-joining method, after which they were swapped using the nearest-neighbor interchange method. Phylogenetic inferences regarding the maximum likelihood trees were made using the PAUP program (Swofford, 2002). The Bayesian tree of full-length genomes was also obtained through the use of the MrBayes program, version 3.1.2 (Ronquist and Huelsenbeck, 2003), with the GTR+ gamma correction model. We made two independent runs of 7 × 10^6 generations each (the initial 10% of the generations were discarded as burn-in) and sampled every 100th generation.

**Analysis of selection pressures**

We used pairwise and maximum likelihood codon-specific methods to estimate selection pressures. Initially, we compiled data sets as previously described and determined the mean dS/dD using the pairwise method of Nei and Gojobori as implemented in the MEGA3 software package, version 3.1 (Kumar, Tamura, and Nei, 2004). For the codon-specific analyses, we used the CODEML program of the PAML package, version 3.14 (Yang, 1998). This approach compares the fit to the data of various models of codon evolution, which differ in the distribution of dS/dD (ω) among sites, and takes into account the phylogenetic relationships among the sequences:

- **Model 0** (one-ratio) assumes a single ω for all sites in the alignment and hence is the simplest model specified.
- **Model 1** (neutral) allows different proportions of conserved sites (ω = 0) and neutral sites (ω = 1), both estimated from the data.
- **Model 2** (selection) adds an additional class of sites, and its ω ratio (which can be > 1) is estimated from the data.
- **Model 3** (discrete) also allows positive selection by incorporating three categories of codon sites, and its ω value at each site is estimated from the data.
- **Model 7** specifies ten categories of ω, none of which may be <1, and the model therefore allows only neutral evolution.
- **Model 8** employs the ten model 7 categories and incorporates an eleventh class of sites that can take on any value of ω, including those supporting positive selection.

Nested models can be compared using a standard likelihood ratio test. Significant evidence for positive selection is provided if model 2 or (as is more common) model 3 significantly rejects the null hypotheses of model 0 and model 1, and if model 8 significantly rejects that of model 7, as well as if the favored models contain a class of codons in which ω is > 1.

**Acknowledgments**

This study received financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Foundation for the Support of Research in the State of São Paulo; grant number: 98/14381-4). We are in debt with Jorge Casseb and Michelle Hendry for the support with the Peptide-binding assay for part of samples, and Antonio Eduardo Beneduto Silvino and Maria Lucia Ferraz for providing the samples collected in 1983.

**References**


