J Mol Evol (2008) 66:598–604
DOI 10.1007/s00239-008-9112-6

Positive Selection on HIV Accessory Proteins and the Analysis of Molecular Adaptation After Interspecies Transmission

André E. R. Soares · Marcelo A. Soares · Carlos G. Schrago

Received: 30 November 2007 / Accepted: 23 April 2008 / Published online: 9 May 2008

Abstract Studies examining positive selection on accessory proteins of HIV are rare, although these proteins play an important role in pathogenesis in vivo. Moreover, despite the biological relevance of analyses of molecular adaptation after viral transmission between species, the issue is still poorly studied. Here we present evidence that accessory proteins are subjected to positive selective forces exclusively in HIV. This scenario suggests that accessory protein genes are under adaptive evolution in HIV clades, while in SIVcpz such a phenomenon could not be detected. As a result, we show that comparative studies are critical to carry out functional investigation of positively selected protein sites, as they might help to achieve a better comprehension of the biology of HIV pathogenesis.

Keywords Human immunodeficiency virus · Primate lentivirus · Accessory protein · Adaptive evolution

Introduction

The human immunodeficiency virus (HIV) is divided into two types, 1 and 2, according to its phylogenetic relatedness to simian immunodeficiency viruses (SIV). HIV-1 is responsible for the majority of worldwide infections and is further divided into three groups of independent origins: M, N, and O (Corbet et al. 2000; Gao et al. 1999; Simon et al. 1998). Groups M and N are descendants of the SIV that circulates in chimpanzees in West Africa (SIVcpz) (Sharp et al. 2005). Group O has been recently associated with SIV found in wild gorillas (Van Heuverswyn et al. 2006). Differently, HIV-2 evolved from the SIV found in sooty mangabeys (SIVsm) (Gao et al. 1992; Hirsch et al. 1989).

The genomes of HIV are composed of nine genes, which encode three structural, two regulatory, and four accessory proteins. The structural genes *gag*, *pol* and *env* are also found in other retroviral lineages, as well as the *tat* and *rev* regulatory genes. The accessory genes of both SIVcpz and HIV-1 consist of *nef*, *vif*, *vpr*, and *vpu*, while in HIV-2 the *vpu* gene is absent and the *vpx* gene is found (Tristem et al. 1992).

The interspecies transmission events that gave rise to HIV epidemics were estimated to have occurred relatively recently (Korber et al. 2000). After crossing the species barrier, genomes of SIV lineages probably underwent adaptive changes to the newly infected human host (Sharp et al. 2005; Wain et al. 2007). These changes may reveal sites relevant to the understanding of the biology of HIV pathogenesis. Moreover, the use of antiretroviral drug therapies also prepared the ground for selection of resistant viral variants. Such scenarios have motivated many investigations of adaptive molecular evolution of HIV genomes, which have focused mainly on structural and regulatory proteins (Choisy et al. 2004; Pan et al. 2007; Wain et al. 2007; Yang et al. 2003).

Studies conducting evolutionary analyses of the accessory proteins Nef, Vif, Vpr, and Vpu are uncommon, although authors have reported that the *nef* gene is under

Electronic supplementary material The online version of this article (doi:10.1007/s00239-008-9112-6) contains supplementary material, which is available to authorized users.

A. E. R. Soares · M. A. Soares · C. G. Schrago (✉)
Departamento de Genética, Universidade Federal do Rio de Janeiro, Ilha do Fundao, Rio de Janeiro, RJ CEP 21.941-590, Brazil
e-mail: guerra@biologia.ufrj.br

M. A. Soares
Divisão de Genética, Instituto Nacional do Câncer, Rio de Janeiro, Brazil

positive selective pressure in HIV-1 (de Oliveira et al. 2004; Yang et al. 2003; Zanotto et al. 1999). Furthermore, only diversifying selection within lineages is typically investigated, while directional selection is rarely considered.

Accessory proteins are unusual because, although they might not be necessary to in vitro replication, they play pivotal roles in pathogenesis in vivo (Le Rouzic and Benichou 2005; Seelamgari et al. 2004), and thus, evolutionary studies of these proteins are potentially interesting. For example, the Vif protein is necessary to escape the host intracellular antiviral factor APOBEC3G (Marin et al. 2003). Nef downregulates the cell surface expression of CD4, disturbs T cell activation, and stimulates HIV infectivity (Stoddart et al. 2003). Therefore, accessory proteins are good candidates for adaptive evolutionary changes.

The present work analyzes positive selection in HIV and SIVcpz genomes in order to unveil adaptive changes that occurred in accessory proteins after interspecies transmissions of SIV to humans (HIV). We report that the HIV lineages studied present sites under positive selection in *nef*, *vif*, *vpr*, *vpu*, and *vpx*. On the other hand, no accessory protein gene in SIVcpz was inferred to have undergone diversifying selection. Our results point to a scenario of ongoing molecular adaptation of accessory proteins after interspecies transmission.

Materials and Methods

Sequences and Alignments

All data used in this work were retrieved from the Los Alamos National Laboratory HIV Sequence Database (LANL; http://www.hiv-web.lanl.gov). Three HIV lineages were studied to investigate differences in selective pressures on accessory proteins after interspecies transmission: HIV-1M, HIV-1O, and HIV-2. The number of sequences used for each gene on the three lineages is reported in Table 1. HIV-1O, HIV-2, and SIVcpz data consisted of the entire number of sequences available at LANL for the genes under study. For HIV-1M, we sampled the most geographically diverse sequences, avoiding circulating recombinant forms. We also calculated pairwise genetic distances to exclude closely related sequences from the data set.

Table 1 Number of sequences analyzed in this study, all retrieved from the Los Alamos National Laboratory HIV database: numbers in parentheses indicate the sum of branch lengths of the tree in substitutions/codon

	nef	*vif*	*vpr*	*vpu*	*vpx*
HIV-1 M	35 (8.4)	39 (5.9)	41 (8.8)	20 (6.6)	na
HIV-1 O	20 (4.8)	41 (5.7)	43 (4.7)	20 (6.0)	na
HIV-2	24 (9.8)	18 (5.3)	12 (4.3)	na	19 (5.9)
SIVcpz	18 (8.0)	11 (6.4)	9 (6.4)	na	na

Note: na, not applicable

Some sequence sets, which were potentially relevant to our study, had to be eliminated from the analyses. SIVsm sequences were discarded because the majority of available sequences in LANL were obtained from a single individual. The number of sequences available for the SIVcpz *vpu* gene is small, and inferences based on this set would not be robust. All sites containing gaps were excluded from analyses and sites that were present in only a few sequences (<40%) were also discarded.

Alignments were conducted for each gene in each lineage independently. Nucleotide sequences were aligned via their translated amino acid sequences in MEGA 3.1 using CLUSTALW (Thompson et al. 1994). All stop codons were eliminated. HIV-1 and SIVcpz site numbering follows the HIV-1 HXB2 reference, while HIV-2 follows SIV SMM239. Alignments of joint samples of HIV-1 M, HIV-1 O, and SIVcpz were also obtained for the *nef*, *vif*, and *vpr* genes for further analyses of selection acting on branches of the phylogeny. In this case, data sets were reduced to 25 sequences for each gene for computer feasibility of the analyses.

Phylogenetic Analysis

Phylogenetic trees of the alignments investigated were inferred by means of the maximum likelihood method using the heuristic search algorithm available in PAUP 4b10 with default settings (Swofford 2003). The model of sequence evolution used in PAUP was estimated by the likelihood ratio test (LRT) in MODELTEST 3.6 (Posada and Crandall 1998) with the significance level set at 1%.

Detection of Recombination Breakpoints

Recombination is known to affect inference of positive selection (Anisimova et al. 2003), and thus, we used the GARD method (Pond et al. 2006) available at the http://www.DataMonkey.org server to search for recombination breakpoints. The HKY85 model was used in GARD analyses and rate variation was implemented via a general discrete distribution with three rate classes (Pond and Frost 2005b).

Test of Synonymous Rate Heterogeneity

Accessory protein genes present partially overlapping reading frames, which results in variation of synonymous rates (Hughes et al. 2001). To make inference of adaptive molecular evolution more conservative, we tested whether models that permit variation of synonymous rates would fit the data better than models that fix synonymous rate at 1

(e.g.,Yang et al. 2000). The test is implemented in the HyPhy package (Pond et al. 2005) by comparing the log-likelihoods of the nonsynonymous and dual models (GDD 3 × 3) via LRT with 4 degrees of freedom (df) and a significance level of 5% (Pond and Muse 2005).

Analysis of Adaptive Molecular Evolution

Analyses of positive selection on data sets were conducted in a maximum likelihood framework in two steps: (i) test of occurrence of positive selection and (ii) analysis of lineage-specific adaptive molecular evolution. The existence of positive selection on gene alignments was investigated by two methodological approaches: the widely used tests available in the PAML 4 package (Yang 2007) and the recently proposed modification of PAML's M1a-M2a test, which incorporates partitioning of data and variation of synonymous substitution rates (Scheffler et al. 2006), available in the HyPhy package.

First, three pairs of nested codon evolutionary models (M1a-M2a, M7-M8, and M8-M8a) were used to test against the null hypothesis of no positive selection (Swanson et al. 2003; Wong et al. 2004; Yang et al. 2000) in the Codeml program of the PAML. In each group, maximum likelihood tree topologies obtained in PAUP were entered in Codeml and log-likelihoods of the one-rate (M0) model were estimated for each alignment. Data sets in which the M1a model was significantly better than M0 were subjected to selection investigation. The null hypothesis of the M1a-M2a and M7-M8 tests were rejected if the likelihood ratio statistic was significant at χ^2_2 ($\alpha = 0.05$). The M8-M8a likelihood ratio statistic was compared with a 50:50 mixture of point mass 0 and χ^2_1 ($\alpha = 0.05$) (Swanson et al. 2003). Samples in which positive selection could be assumed by the three tests were considered for further analysis of adaptive molecular evolution at protein sites. A site was admitted as positively selected when the posterior probability of a codon in belonging to the $\omega > 1$ class was >95% using the Bayes Empirical Bayes approach (Yang et al. 2005). In Codeml, we considered to be under adaptive molecular evolution the consensus of all three model comparisons.

The second test of positive selection was implemented by the PARRIS algorithm applied to the results of the GARD analyses to permit partitioning of data, which controls the rate of false positives (Scheffler et al. 2006). The MG94 × HKY85 model of codon evolution was used. Heterogeneity of synonymous rates along sequences was implemented by using a dual general discrete distribution (GDD) with three bins each for synonymous and nonsynonymous rates (Pond and Muse 2005). The null hypothesis of no $\omega > 1$ class was rejected if the likelihood ratio statistic was significant at $\alpha = 0.05$. Positively selected sites were identified at a posterior probability cutoff of 95%. Finally, adaptive molecular evolution on branches of the phylogeny (item ii above) was investigated by means of the genetic algorithm of Pond and Frost (2005a) implemented in the GABranch application of the DataMonkey.org server, also using the MG94 × HKY85 model.

Results

Recombination analyses in GARD detected single breakpoints in several gene alignments. In the *nef* gene, HIV-1M presented a breakpoint at nucleotide site 202; in HIV-1O, a breakpoint was detected in *nef* site 124; in HIV-2, at site 450; and in SIVcpz at site 186. Recombination was also inferred in *vif* of HIV-2 (nucleotide site 479) and SIVcpz (72), as well as in *vpx* of HIV-2 (521). No breakpoints were found in *vpr* and *vpu*. All tests of synonymous rate variation (nonsynonymous vs dual GDD 3 × 3 models) along codon sites were significant at $p < 0.01$. All M0-M1a comparisons rejected the null hypothesis of only one ω rate along codons in accessory protein genes ($p < 0.01$).

The *nef* gene of both HIV-1 lineages presented sites under positive selection in Codeml and HyPhy (Table 2). The M2a and M8 models of codon evolution fit the data significantly better than the M1a and M7 models, respectively, as well as the M8–M8a test ($p < 0.01$ for HIV-1M and HIV-1O). HIV-2 *nef* was also inferred to have positively selected sites in both Codeml and HyPhy. Sites that underwent adaptive molecular evolution could not be found in the *nef* gene of SIVcpz with either methodological approach, since models M1a and M7 presented the same log-likelihoods as models M2a and M8, respectively. Therefore, models M2a and M8 were essentially reduced to

Table 2 Likelihood ratio statistics ($2\Delta\ell$) for M1a-M2a, M7-M8, M8-M8a, and PARRIS tests of accessory protein genes in each lineage

	nef	*vif*	*vpr*	*vpu*	*vpx*
HIV-1 M	49.7, 51.1, 33.6, 30.8[a]	28.4, 32.0, 25.2, 8.6	18.4, 31.7, 16.4, 11.2	40.7, 43.5, 36.5, 34.9	na
HIV-1 O	37.7, 36.8, 33.5, 15.9	138.1, 133.4, 122.0, 53.6	41.7, 40.7, 36.7, 7.5	0.0, 2.1, 0.9, 2.2	na
HIV-2	42.6, 884.2, 911.1, 128.4	7.0, 12.7, 9.5, 3.9	7.6, 5.1, 4.9, 10.6	na	13.8, 26.3, 17.4, 7.2
SIVcpz	0.0, 0.0, 0.0, 0.0	0.0, 2.4, 1.5, 0.0	0.0, 0.6, 0.0, 0.1	na	na

Note: na, not applicable. Values underlined are significant at $p < 0.05$

[a] M1a-M2a, M7-M8, M8-M8a, and PARRIS $2\Delta\ell$, respectively

models M1a and M7 with the positive selection category ω close to one. The log-likelihood of the modified M2a model in PARRIS was also identical to the null model, thus, no positive selection was inferred.

A similar pattern was found in the *vif* gene. HIV-1M and HIV-1O had significantly higher log-likelihoods under models M2a and M8 of Codeml and model M2a of PARRIS ($p < 0.01$) (Table 2). The positive selection models also explained data better than Codeml's models M1a and M7 in the HIV-2 data set. In HyPhy, however, the modified M1a-M2a comparison was not significant in HIV-2. In the SIVcpz sample, no statistical evidence of positive selection was found by any test. In Codeml, log-likelihoods of the M1a and M2a models were identical and the LRT was not significant for the M7-M8 and M8-M8a comparisons. The same occurred in HyPhy's PARRIS.

In the *vpr* gene, adaptive molecular evolution could be detected on HIV-1M and HIV-1O by Codeml and HyPhy. Both samples presented significant differences in log-likelihoods between model M1a and model M2a ($p < 0.01$). In Codeml, the comparison of models M1a-M2a and M8-M8a in HIV-2 were significant, although the log-likelihood of model M8 was not significantly higher than that of M7. In PARRIS, the null hypothesis of no positive selection in HIV-2 could be rejected ($p < 0.01$). Again, SIVcpz did not present sites under positive Darwinian selection in either Codeml or HyPhy. Models M1a and M7 could not be discarded in favor of M2a or M8, since log-likelihoods were almost identical.

The *vpu* gene of HIV-1M was subjected to positive selection in all model comparisons independent of the method used. However, there was no statistical evidence for sites in the $\omega > 1$ category in HIV-1O in either algorithm. The *vpx* gene, which is exclusive of HIV-2, was inferred to presented sites belonging to the positive selection category of models M2a and M8 of Codeml and modified model M2a of HyPhy ($p < 0.01$).

The consensus sites of Codeml analyses reveal that the Nef protein of HIV-1 M has five sites detected as being under positive selection with posterior probabilities ≥95% (sites 11, 22, 130, 148, and 185), while one site was detected in HIV-1O (site) and two sites in HIV-2 (sites 32 and 75) (Table 3). The Vif protein of HIV-1M has two sites under positive selection (sites 31 and 39), while HIV-1O has six sites (sites 35, 39, 46, 61, 63, and 136). In HIV-2 only one site was detected under positive selection (site 107). HIV-1M and HIV-1O *vif* share the same codon 39 under positive selection. *vpr* presents two positively selected sites in HIV-1M (sites 84 and 86) and four sites in HIV-1O (sites 23, 37, 41, and 72). No positively selected codon was found in Codeml consensus for HIV-2 and SIVcpz. HIV-1M and HIV-1O sites were not coincident. HIV-1M was the only HIV lineage to present codons under positive selection in the *vpu* gene (sites 2, 3, 5, and 7). Finally, Codeml consensus of HIV-2 sites resulted in only one codon of *vpx* under positive selection (site 64).

In HyPhy, the PARRIS method detected four sites under positive selection in the HIV-1M Nef protein (sites 11, 22, 130, and 185), with strikingly similar results to Codeml (Table 4). On the other hand, in the HIV-1M Vif protein analysis, nine sites (nos. 31, 36, 37, 39, 47, 63, 155, 159, and 167) were found to be under positive selection, and those

Table 3 Positively selected sites on HIV and SIV accessory proteins; consensus of all model comparisons performed in Codeml

	nef	*vif*	*vpr*	*vpu*	*vpx*
HIV-1 M	**11**, **22**, **130**, 148, **185**	**31**, **39**	**84**, **86**	**2**, 3, **5**, **7**	na
HIV-1 O	**170**	35, **39**, **46**, **61**, **63**, **136**	**22**, **37**, **41**, **7 2**	–	na
HIV-2	32,75	107	–	na	**64**
SIVcpz	–	–	–	na	na

Note: na, not applicable. Sites in boldface have posterior probabilities ≥99%; otherwise, 95% ≤ p < 99%. Site numbering follows reference sequences from LANL: HXB2CG (accession no. K03455) for HIV-1 and SMM239 (accession no. M33262) for HIV-2

Table 4 Positively selected sites on HIV and SIV accessory proteins obtained in HyPhy software, using the PARRIS method with rate variation modeled by a 3 × 3 GDD, MG94xHKY85

	nef	*vif*	*vpr*	*vpu*	*vpx*
HIV–1 M	11, 22, 130, 185	31, 36, 37, 39, 47, 63, 155, 159, 167	28, 84, 86	2, 5, 7	na
HIV–1 O	24, 28, 170	20, 37, 39, 46, 61, 63	36	–	na
HIV–2	31, 33, 75, 81, 110, 182, 222, 227, 228	–	83	na	64
SIVcpz	–	–	–	na	na

Note: na, not applicable. Site numbering follows reference sequences from LANL: HXB2CG (accession no. K03455) for HIV-1 and SMM239 (accession no. M33262) or HIV-2

Table 5 Consensus of Codeml and HyPhy positively selected sites

	nef	*vif*	*vpr*	*vpu*	*vpx*
HIV-1 M	11, 22, 130, 185	31, 39	28, 84, 86	2, 5, 7	na
HIV-1 O	170	39, 46, 61, 63	–	–	na
HIV-2	75	–	–	na	64
SIVcpz	–	–	–	na	na

Note: na, not applicable

were not duplicated in Codeml. The HIV-1M Vpr protein yielded three positively selected codons when recombination and synonymous rate heterogeneity were considered (nos. 24, 84, and 86), versus two sites in Codeml (nos. 84 and 86). The Vpu protein had three sites detected as belonging to the $\omega > 1$ class (sites 2, 5, and 7), while Codeml additionally detected codon position 3. Sites 24, 28, and 170 were detected in the HIV-1O Nef protein, while Vif had six sites under positive selection (sites 20, 37, 39, 46, 61, and 63). Only one site was detected to be under positive selection in HIV-1O Vpr (site 36) and no site was inferred in the *vpu* gene. HIV-2 had several sites detected by the PARRIS method (sites 31, 33, 75, 81, 110, 182, 222, 227, and 228), while Codeml detected only two sites in this same protein. No codon was inferred as positively selected in the Vif protein, and only one site was detected in the HIV-2 Vpr (no. 83) and Vpx (no. 64) proteins.

The consensus codon sites of the two methodological approaches used resulted in several matches (Table 5). In *nef*, codons 11, 22, 130, and 185 in HIV-1M, site 170 in HIV-1O, and codon 75 in HIV-2 were inferred under adaptive molecular evolution in all analyses conducted. In *vif*, HIV-1M (sites 31 and 39) and HIV1-O (sites 39, 46, 61, and 63) presented corresponding sites in both methods applied. These were also found in the HIV-1M *vpr* (sites 28, 84, and 86) and *vpu* (sites 2, 5, and 7) genes, as well as in the *vpu* gene of HIV-2 (site 64).

The GABranch analysis, which searched for episodes of adaptive evolution along the phylogeny of the lineages, revealed that several branches could be assigned to d_N/d_S values > 1 (see Supplementary Material). In *nef*, positive selection episodes ($d_N/d_S = 2.0$) were found within the HIV-1M, HIV-1O, and SIVcpz clades, but not between lineages. A similar result was found in *vif*, but with a lower inferred d_N/d_S value (1.2). In *vpr*, however, the genetic algorithm could not associate any branch with $d_N/d_S > 1$, since such class was not estimated from the data.

Discussion

In contrast with earlier works (Yang et al. 2003; Zanotto et al. 1999), we broadened the study of accessory proteins analyzing various HIV subtypes, including sequences from a worldwide perspective to avoid patterns of evolution from a single population. We also separated each major group of HIV, considering their different phylogenetic histories, and compared them to SIVcpz to shed light on the evolution of those proteins. Additionally, we applied a test of positive selection that deals with issues relevant to the genes studied, such as data partitioning (caused by recombination) and heterogeneous synonymous rates along genes.

Our results show the absence of positive selection in SIVcpz, while it occurs in all lineages of HIV. Since SIVcpz is closely related to HIV-1 and the chimpanzee (*Pan troglodytes troglodytes*) is the natural reservoir of HIV-1 (Keele et al. 2006), we would expect SIVcpz to be under the same selective regime as HIV, but this was not the case. Other authors have also pointed out that the simian viruses that were transmitted to humans by cross-species events experienced a series of changes that helped HIV-1 and 2 to successfully infect the new host (Heeney et al. 2006).

The absence of positive selection on samples of SIVcpz accessory proteins may have biological and technical explanations. Biologically, differences between SIV and HIV are also found in the structuring of the genetic variation in natural populations. Moreover, the SIVcpz infection is asymptomatic in chimpanzees, probably due to a long period of exposure of the chimpanzee populations to the virus, which led to an equilibrium state between the virus and the host.

Technically, the small sample sizes of SIVcpz data sets could reduce the power of selection analyses. However, the sums of branch lengths of SIVcpz gene trees (see Table 1) are adequate to such investigations as indicated by simulation studies (Anisimova et al. 2001; Anisimova et al. 2002). Nevertheless, in order to check the power of the tests, all HIV analyses were repeated utilizing reduced samples sizes of 10 sequences. This task was performed to verify reproducibility of results. Even with reduced sample sizes, HIV genes under positive selection maintained similar patterns, with significant log-likelihood differences between M1a-M2a, M7-M8, and M8-M8a Codeml models (Table 6). In PARRIS, however, small samples have indeed reduced the power of the tests.

We noticed that when samples were reduced, the GARD algorithm found a significantly higher number of recombination breakpoints. For example, in the *nef* gene of HIV-2, when 24 sequences were used, only one breakpoint was inferred. When downsized to 10 sequences, four breakpoints were estimated. We suppose that this has greatly reduced the power of the PARRIS test, which could not reject the null hypothesis of no positive selection category in small data sets. Actually, this fact has already been reported by Scheffler et al. (2006). We believe that PARRIS results for the small data sets do not invalidate our conclusions because the SIVcpz samples used presented

Table 6 Likelihood ratio statistics ($2\Delta\ell$) for M1a-M2a, M7-M8, M8-M8a, and PARRIS test of reduced samples (10 sequences) of accessory protein genes in each lineage

	nef	*vif*	*vpr*	*Vpu*	*vpx*
HIV-1 M	7.5, 9.2, 5.6, 1.3[a]	2.6, 6.9, 3.0, 2.8	8.66, 10.6, 9.4, 9.0	17.1, 25.6, 18.0, 5.8	na
HIV-1 O	16.9, 17.3, 16.8, 13.8	22.9, 23.8, 16.2, 15.4	6.2, 9.6, 7.9, 1.7	6.0, 8.4, 6.9, 3.0	na
HIV-2	32.1, 40.6, 34.3, 1.7	17.3, 17.0, 15.7, 0.0	0.7, 1.3, 0.8, 0.5	na	3.7, 10.4, 5.4, 2.4

Note: na, not applicable. Values underlined are significant at $p < 0.05$

[a] M1a-M2a, M7-M8, M8-M8a, and PARRIS $2\Delta\ell$, respectively

adequate levels of diversity. The LRT rejected the null hypothesis in samples where the number of sequences and the sum of branch lengths were lower than in SIVcpz. For example, the HIV-2 *vpx* sample was composed of 19 sequences and a tree length of 5.9 and presented a significant likelihood ratio statistic, while the SIVcpz *nef* set contained 20 sequences and one of the highest tree lengths (8.0); nevertheless, all methods were unable to reject the null hypothesis of no positive selection.

The GABranch method could not assert basal branches of the HIV and SIVcpz lineages to $d_N/d_S > 1$ categories. Therefore, directional selection leading to adaptive changes after interspecies transmission was not corroborated. Hence, the most likely explanation for the sites under positive selection within HIV lineages is diversifying selection on proteins. For instance, the positively selected sites inferred may account for different responses of the immunodeficiency virus to the immune system of their hosts, especially the cytotoxic T lymphocyte (CTL) response.

Some of the sites identified play roles in the HIV life cycle that were already described. For example, the N-terminal portion of HIV Vif (sites 1–64) binds to the viral RNA. This bound can be undone by interaction with Gag protein and is supposedly a mechanism of interaction between the viral RNA and Gag and, also, a way to maintain its correct folding and preserve it from cell inhibitors (Zhang et al. 2000). It is also essential to the packaging of Vif into virions, as part of the viral nucleoprotein complex (Khan et al. 2001). We found several sites in this part of the protein to be under positive selection (sites 36, 46, 61, and 63). Our analyses also confirmed positive selective pressures in the N-terminal (transmembrane) domain of the HIV-1M Vpu protein, which have been correlated with the efficiency of virus particle release from the plasma membrane of infected cells (Schubert et al. 1996).

It is noteworthy that some codons from HIV accessory genes inferred by our study have also been evidenced by Yang et al. (2003), such as codons 31 and 39 in HIV-1M *vif* and codon 84 in HIV-1M *vpr*. The consistency of the results from our study and that of Yang and collaborators strengthens the reliability of our estimates.

As stated previously, although accessory proteins are not structural, they interact with several cellular pathways. Substitutions in these genes, particularly in sites under positive selection, might be governing keys to viral infection and onset of disease in different species. This reason reinforces the need to understand accessory proteins under a comparative scenario, as their evolution and response to selective pressures are probably unique from the HIV perspective. The knowledge of the differences between immune system responses in both HIV and SIV hosts might lead us to a better understanding of the success of viral pathogenesis of HIV in humans and why chimpanzees do not develop AIDS.

Acknowledgments This work is part of the requirements for the Master's degree in Genetics of A.E.R.S., who is supported by a scholarship from the Brazilian National Research Council (CNPq). M.A.S. is supported by CNPq Grant 403589/2004-5 and Rio de Janeiro Science Foundation (FAPERJ) Grant E-26/170-545/2004. C.G.S. is supported by FAPERJ Grants 171.157/2006 and 110.627/2007 and grants from the FAPERJ programs Pensa Rio 2007 and Apoio às Instituições de Pesquisa 2007 to Claudia A. M. Russo. We would also like to thank the reviewers and the editor for the insightful comments that greatly enriched our article.

References

Anisimova M, Bielawski JP, Yang ZH (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Anisimova M, Bielawski JP, Yang ZH (2002) Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19:950–958

Anisimova M, Nielsen R, Yang ZH (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Choisy M, Woelk CH, Guegan JF, Robertson DL (2004) Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes. J Virol 78:1962–1970

Corbet S, Muller-Trutwin MC, Versmisse P, Delarue S, Ayouba A, Lewis J, Brunak S, Martin P, Brun-Vezinet F, Simon F, Barre-Sinoussi F, Mauclere P (2000) env Sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. J Virol 74:529–534

de Oliveira T, Salemi M, Gordon M, Vandamme AM, van Rensburg E, Engelbrecht S, Coovadia HM, Cassol S (2004) Mapping sites

of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design? Genetics 167:1047–1058

Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, Greene BM, Sharp PM, Shaw GM, Hahn BH (1992) Human infection by genetically diverse Sivsm-related HIV-2 in West Africa. Nature 358:495–499

Gao F, Bailes E, Robertson DL, Chen YL, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397:436–441

Heeney JL, Dalgleish AG, Weiss RA (2006) Origins of HIV and the evolution of resistance to AIDS. Science 313:462–466

Hirsch VM, Olmsted RA, Murpheycorb M, Purcell RH, Johnson PR (1989) An African primate lentivirus (Sivsm) closely related to HIV-2. Nature 339:389–392

Hughes AL, Westover K, Da Silva J, O'Connor DH, Watkins DI (2001) Simultaneous positive and purifying selection on overlapping reading frames of the tat and vpr genes of simian immunodeficiency virus. J Virol 75:7966–7972

Keele BF, Van Heuverswyn F, Li YY, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen YL, Wain LV, Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JFY, Sharp PM, Shaw GM, Peeters M, Hahn BH (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313:523–526

Khan MA, Aberham C, Kao S, Akari H, Gorelick R, Bour S, Strebel K (2001) Human immunodeficiency virus type I Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. J Virol 75:7252–7265

Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T (2000) Timing the ancestor of the HIV-1 pandemic strains. Science 288:1789–1796

Le Rouzic E, Benichou S (2005) The Vpr protein from HIV-1: distinct roles along the viral life cycle. Retrovirology 2:11. Available at: http://www.retrovirology.com/content/2/1/11

Marin M, Rose KM, Kozak SL, Kabat D (2003) HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nature Med 9:1398–1403

Pan C, Kim J, Chen LM, Wang Q, Lee C (2007) The HIV positive selection mutation database. Nucleic Acids Res 35:D371–D375

Pond SLK, Frost SDW (2005a) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol 22:478–485

Pond SLK, Frost SDW (2005b) A simple hierarchical approach to modeling distributions of substitution rates. Mol Biol Evol 22:223–234

Pond SK, Muse SV (2005) Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22:2375–2385

Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Scheffler K, Martin DP, Seoighe C (2006) Robust inference of positive selection from recombining coding sequences. Bioinformatics 22:2493–2499

Schubert U, FerrerMontiel AV, OblattMontal M, Henklein P, Strebel K, Montal M (1996) Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. FEBS Lett 398:12–18

Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng LW, Dadgar S, Bottazzi ME, Ghedin E, Pumfery A, Kashanchi F (2004) Role of viral regulatory and accessory proteins in HIV-1 replication. Frontiers Biosci 9:2388–2413

Sharp PM, Shaw GM, Hahn BH (2005) Simian immunodeficiency virus infection of chimpanzees. J Virol 79:3891–3902

Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barre-Sinoussi F, Brun-Vezinet F (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nature Med 4:1032–1037

Stoddart CA, Geleziunas R, Ferrell S, Linquist-Stepps V, Moreno ME, Bare C, Xu WD, Yonemoto W, Bresnahan PA, McCune JM, Greene WC (2003) Human immunodeficiency virus type 1 Nef-mediated downregulation of CD4 correlates with Nef enhancement of viral pathogenesis. J Virol 77:2124–2133

Swanson WJ, Nielsen R, Yang QF (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20

Swofford DL (2003) PAUP* 4b10. Phylogenetic analysis using parsymony (*and other methods). Sinauer Associates, Sunderland, MA

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tristem M, Marshall C, Karpas A, Hill F (1992) Evolution of the primate lentiviruses—evidence from Vpx and Vpr. EMBO J 11:3405–3412

Van Heuverswyn F, Li YY, Neel C, Bailes E, Keele BF, Liu WM, Loul S, Butel C, Liegeois F, Bienvenue Y, Ngolle EM, Sharp PM, Shaw GM, Delaporte E, Hahn BH, Peeters M (2006) Human immunodeficiency viruses—SIV infection in wild gorillas. Nature 444:164–164

Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, Li YY, Takehisa J, Ngole EM, Shaw GM, Peeters M, Hahn BH, Sharp PM (2007) Adaptation of HIV-1 to its human host. Mol Biol Evol 24:1853–1860

Wong WSW, Yang ZH, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051

Yang W, Bielawski JP, Yang ZH (2003) Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. J Mol Evol 57:212–221

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Zanotto PMD, Kallas EG, de Souza RF, Holmes EC (1999) Genealogical evidence for positive selection in the nef gene of HIV-1. Genetics 153:1077–1089

Zhang H, Pomerantz RJ, Dornadula G, Sun Y (2000) Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. J Virol 74:8252–8261