MDPI

Site-Specific Evolutionary Rate Shifts in HIV-1 and SIV

Maoz Gelbart and **Adi Stern** *

The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences,
Tel Aviv University, Tel Aviv 6997801, Israel; maozgelbart@mail.tau.ac.il
* Correspondence: sternadi@tauex.tau.ac.il

check for updates

Abstract: Site-specific evolutionary rate shifts are defined as protein sites, where the rate of substitution has changed dramatically across the phylogeny. With respect to a given clade, sites may either undergo a rate acceleration or a rate deceleration, reflecting a site that was conserved and became variable, or vice-versa, respectively. Sites displaying such a dramatic evolutionary change may point to a loss or gain of function at the protein site, reflecting adaptation, or they may indicate epistatic interactions among sites. Here, we analyzed full genomes of HIV and SIV-1 and identified 271 rate-shifting sites along the HIV-1/SIV phylogeny. The majority of rate shifts occurred at long branches, often corresponding to cross-species transmission branches. We noted that in most proteins, the number of rate accelerations and decelerations was equal, and we suggest that this reflects epistatic interactions among sites. However, several accessory proteins were enriched for either accelerations or decelerations, and we suggest that this may be a signature of adaptation to new hosts. Interestingly, the non-pandemic HIV-1 group O clade exhibited a substantially higher number of rate-shift events than the pandemic group M clade. We propose that this may be a reflection of the height of the species barrier between gorillas and humans versus chimpanzees and humans. Our results provide a genome-wide view of the constraints operating on proteins of HIV-1 and SIV.

Keywords: phylogenetics; rate shifts; cross-species transmission; HIV

1. Introduction

The human immunodeficiency viruses HIV-1 and HIV-2 are the causative agents of AIDS in humans, infecting millions of people worldwide. Both viruses emerged from a clade of lentiviruses known as the simian immunodeficiency viruses (SIV), which naturally infect a variety of non-human primate species. HIV in humans arose from several independent transmission events of primate SIVs that resulted in HIV-1 groups M and N (from SIV infecting chimpanzees, SIVcpz), HIV-1 groups O and P (from SIV infecting gorillas, SIVgor), and HIV-2 groups A through H (from SIV naturally infecting sooty mangabeys, SIVsmm) [1–3]. The gorilla-infecting lentivirus, SIVgor, is itself a result of a transmission of SIVcpz from chimpanzees to gorillas [4,5]. Phylogenetic analyses date the most common recent ancestors of HIV groups M and O to the beginning of the 20th century, making it a relatively new human pathogen [6,7]. Similar analyses of SIVgor date the most recent common ancestor of this virus in gorilla somewhere in the 19th century [4]. SIVcpz itself was found to be a transmission from other primates, leading to two lineages of SIVcpz: SIVcpz*ptt* infecting the chimpanzee subspecies *Pan troglodytes troglodytes* of central Africa, and SIVcpz*pts* infecting the *Pan troglodytes schweinfurthii* chimpanzee subspecies of eastern Africa [5,8–12].

How viruses are able to cross the species barrier is a subject of much interest, since many pandemic human viruses arose from zoonosis events: the influenza strain of Spanish flu H1N1, Ebola Virus,

SARS, and SARS-CoV-2 coronaviruses [13–15]. Due to differences between species, virus adaptation to new host species occurs at multiple stages in the virus infection cycle. To date, studies have found that the Gag, Env, and in particular, the Vpu protein of HIV likely underwent adaptation during the transition to a human host [16–23].

Protein adaptations are reflected in the history of genomes and may be manifested as changes in the rate of evolution at the adapting sites, known as rate shifts. For example, if a site was under weak constraint in the past (and exhibited a fast rate of evolution), following acquisition of a new function, the site will be under a higher degree of constraint (and will exhibit slower rate of evolution) [24–27]. However, recent work has shown that pervasive epistatic interactions among protein sites may also lead to observed changes at the rate of evolution at a site [28,29]. Accordingly, the presence of a certain amino acid at one site may lead to the tolerance of various amino acids at a second site and, hence, potentially a high rate of evolution at this second site. However, a substitution at the first site to another amino-acid may dictate loss of tolerance at the second site, also known as entrenchment, and hence, a lower rate of evolution [30–33]. Such epistatic patterns have indeed been found for the Env protein of HIV-1 [34,35].

Until today, the lack of SIVcpz and group O full genomic sequences limited the ability to infer rate shifts across cross-species transmission events. Here, we utilized the growing availability of diverse and full HIV-1 and SIV genomes to identify sites in SIVcpz/SIVgor/HIV-1 clades whose evolutionary rate significantly changed across clades. We next attempted to characterize the driving force behind the pattern of rate shifts and to infer whether adaptation or epistatic interactions characterize protein evolution across the HIV-1/SIV phylogeny.

2. Methods

In order to collect sequences for this study, the Los Alamos HIV sequence database (available online at http://www.hiv.lanl.gov [36], retrieved at 17 April 2016) was queried for HIV-1 sequences from the same strain that spanned all nine HIV-1 open reading frames (ORFs: *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef*) and for SIVcpz and SIVgor strains that spanned the corresponding ORFs (with the exception of *vpx*). This led to 2004 sequences of HIV-1 group M, 45 sequences of HIV-1 group O, 9 sequences of HIV-1 group N, 2 sequences of HIV-1 group P, 4 sequences of SIVgor, and 29 sequences of SIVcpz. The final study dataset was constructed with an equal number of group M and group O sequences (N = 45), by sampling sequences so that the n most distant strains (in terms of genetic distance) were sampled. Due to extremely high similarity, HIV-1 groups N and P sequences were reduced to a single representative strain from each. The IIIB_LAI strain was added manually as a reference sequence.

Initial MSA's of the nine proteins were performed using PRANK and iteratively improved until convergence [37]. In order to reconstruct of the phylogenetic relationship between the sequences, we concatenated the alignments of Gag, Pol, Vif, Vpr, Tat, and Env and provided this as input for PhyML ([38], File S1). We next used the reconstructed phylogeny as a guide tree to realign each ORF with PRANK (File S1). JpHMM was used to validate that the strains used in the analysis are not inter-group recombinants [39].

In order to identify evolutionary rate shifts, we used RASER [25] to analyze each of the nine proteins separately, with the proteome-based phylogeny as input. RASER is a likelihood-based phylogenetic method for detecting a change in the substitution rates in each site of a given protein, along all possible branches in the phylogeny. First, a likelihood ratio test against a null model of no rate shifts is performed in order to assess if a model enabling rate shifts better fits the data. Next, the posterior probability of rate shift is calculated at each site and sites with a posterior probability higher than 0.6 are reported here. Finally, for each such site, the method lists the lineages where the rate shift occurred with the highest probability and further categorizes each sites as undergoing either a rate deceleration or a rate acceleration. All sites are reported in File S2.

Rate-shift events along the HIV/SIV phylogeny were constructed as the percentage of total rate-shift events at each branch, divided by protein size (Table 1). Another table containing the same rate shifts as percentage from total rate shifts at each branch, without taking into account the protein size, is present as Table S2.

Table 1. The proportion of rate shifts (rate decelerations and rate accelerations in upper panel and lower panel, respectively) detected in each of the major branches of the phylogeny, relative to the protein size. The right-most column shows the absolute number of rate shifts detected in the branch. Cell background color intensity increases with the proportion of rate shifts.

DEC Branch/Protein	Structural Proteins			Regulatory Proteins		Accessory Proteins				Total Number of Rate Shifts
	Env	Gag	Pol	Rev	Tat	Nef	Vif	Vpr	Vpu	
Group M	1%	1%	1%	1%	0%	0%	0%	0%	5%	28
Group O	2%	1%	0%	8%	0%	1%	1%	1%	4%	46
SIVgor + P + O	0%	0%	0%	0%	0%	2%	1%	0%	0%	14
SIVcpz*pts*	1%	1%	0%	0%	0%	3%	2%	1%	5%	27
ACC Branch/Protein	Env	Gag	Pol	Rev	Tat	Nef	Vif	Vpr	Vpu	Total Number of Rate Shifts
Group M	1%	1%	0%	0%	1%	1%	1%	1%	1%	23
Group O	3%	2%	1%	3%	0%	1%	2%	0%	1%	51
SIVgor + P + O	0%	0%	0%	0%	0%	2%	0%	1%	1%	10
SIVcpz*pts*	1%	1%	0%	0%	0%	3%	1%	2%	5%	31

DEC = rate decelerations; ACC = rate accelerations.

3. Results and Discussion

3.1. Finding Evolutionary Rate Changes

We have previously developed a maximum-likelihood-based method called RASER that identifies sites that undergo a significant change in their rate of evolution along a given branch of a phylogenetic tree (see for example Figure 1) [25]. RASER is based on two stages: first, it tests if a rate-shifting model better fits the data than a non-rate-shifting model based on a likelihood ratio test. Second, RASER reports the posterior probability that a protein site evolves under to a rate-shifting model and infers the branch in the phylogeny where the rate shift most likely occurred. Sites are said to be "rate decelerating" or "rate accelerating" at a given branch, if their evolutionary rate is, respectively, slower or faster in one clade of the tree as compared to the complementary clade spanned by the branch. For example, given the branch that separates HIV group M from SIVcpz, a site may be conserved (slower evolutionary rate) in the HIV group M clade and variable (faster evolutionary rate) in the rest of the tree. Notably, these terms are always relative: a rate deceleration at one clade can also be interpreted as rate acceleration in the complementary clade. Here, we sought to identify and characterize rate-shifting sites in the HIV/SIV phylogeny. We queried the Los Alamos HIV database [36] for all available HIV-1, SIVcpz and SIVgor full genome sequences. Rate-shift analyses were performed on the nine translated open reading frames of the virus, and likelihood ratio tests were found to be highly significant in favor of a rate-shift model across all HIV proteins (Table S1).

3.2. Most Rate-Shift Events Are Identified in Long Branches

RASER identified a total of 271 rate-shifting sites along the HIV-1/SIV phylogeny. One hundred and thirty-seven of them were identified as rate deceleration events, and an additional 134 were identified as rate accelerations. Of the 271 rate-shifting sites, 230 were attributed to four branches:

the branch separating group M from SIVcpz, the branch separating group O from SIVgor, the branch separating SIVgor and group O from SIVcpz, and the branch separating SIVcpz*pts* from other viruses (Figure 2B). Notably, these branches are the longest branches in the phylogeny and at least two of them represent cross-species transmission events.
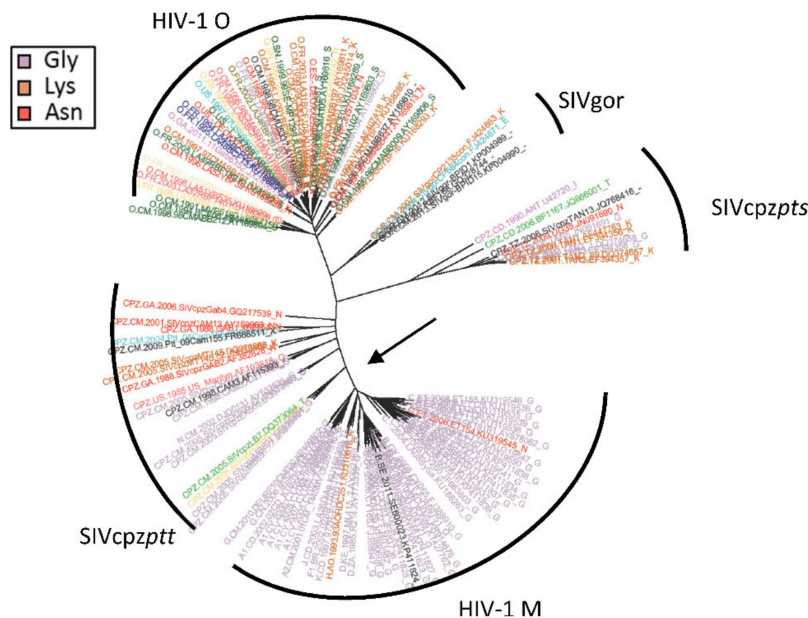

Figure 1. Projection of Env$_{458}$ on the HIV-1/SIV phylogeny. Each leaf (corresponding to an HIV/SIV strain) is color-coded based on the amino acid present at Env$_{458}$. The legend shows the three most prevalent amino acids found at this site. The substitution rate of Env$_{458}$, which interacts with the CD4 receptor, was found to be slower in HIV-1 group M than in the rest of the phylogeny, with the branch separating both clades marked by an arrow.

In order to test for sequence sampling effects on the identified rate-shift patterns, we repeated the analysis with increased amount of group M sequences ($n = 183$), reduced amount of group O sequences ($n = 13$), and no SIVgor sequences (File S3). With this dataset, more rate decelerations were identified in the branch leading to group M, but chi-squared tests for differences in group M rate-shift distributions showed no significant difference ($p = 0.11$ and 0.87 for rate decelerations and accelerations, respectively), indicating that the patterns of rate shifts between the large group M sample and the smaller group M sample are similar.

3.3. Different Rate-Shift Patterns Identified in Different Branches

We calculated the relative contribution of rate-shift events in each protein when controlling for protein size, for each branch corresponding to a virus speciation event (Figure 2B). While rate-shift patterns differed between the different lineages, one common theme was that, at all prominent branches, a large fraction of rate-shift events occurred at non-structural proteins, i.e., the regulatory and accessory proteins of the virus (Table 1 and Table S2). When comparing rate accelerations with the corresponding rate decelerations, the branch separating SIVcpz*pts* from the rest of the tree showed a very similar pattern of rate accelerations/decelerations: a very similar number of rate decelerations and rate accelerations were found in each protein (Figure 2A). A similar pattern was observed in the structural genes: a similar number of sites were inferred as rate accelerations and as rate decelerations in the Gag, Pol, and Env proteins (Table 1).

The pattern of equal accelerations and decelerations led us to conjecture that epistatic interactions coupled with genetic drift may be the most likely explanation for the majority of observed rate shifts. This is based on the fact that (a) a priori, we do not expect any adaptation in the branch separating

SIVcpz*pts* from other viruses, and thus this branch serves as a reference for what to expect under neutral or almost neutral evolution and (b) strong epistatic effects have already been reported in Env [35], which indeed showed large and more or less equal numbers of accelerations and decelerations in most branches (apart from group O).
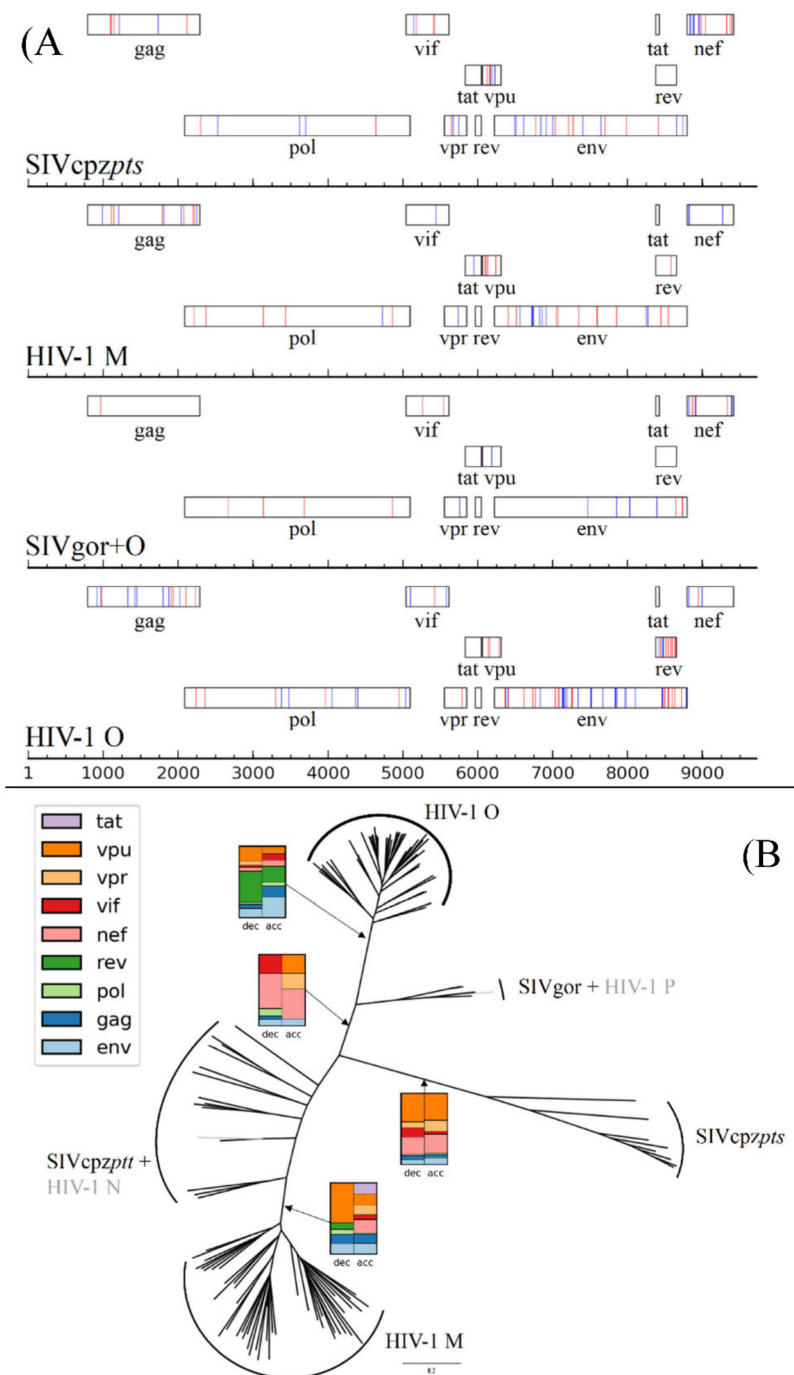

Figure 2. (**A**) Proposed rate deceleration (red) and acceleration (blue) events in prominent branches shown along the genome using the coordinates of the HXB2 reference strain. (**B**) Proposed rate-shift patterns in prominent branches, shown on the phylogeny. Rate shifts (deceleration left and acceleration right) are shown as the percentage of total rate-shift events at each branch, while controlling for protein size.

After establishing a pattern that seemed to be most likely neutral evolution, we went on to focus on outliers in our data: specific proteins where the number of rate accelerations and rate decelerations were substantially different. We noted higher proportions of rate decelerations as compared to rate accelerations in the Vpu protein, in the lineages leading to group M and group O, and in the O-Rev protein. Conversely, we noted more rate accelerations in Nef in the lineage leading to group M. Finally, we observed slightly higher proportions of rate decelerations in Vif and accelerations in Vpr in the lineage leading to SIVgor and group O.

3.4. Rate Shifts and Innate Immunity

We were intrigued by the fact that most of the imbalances found in rate accelerations/decelerations occurred in accessory proteins, which are known to play a role in the response of the virus to the innate immune response, and we survey some of our findings on rate shifts in the context of the interaction between innate immune proteins and HIV proteins.

One such well-characterized interaction is demonstrated by the restriction factor Tetherin, a fast evolving protein that differs between chimpanzees, gorillas, and humans [40,41]. In humans, a significant deletion in the cytoplasmic tail of Tetherin rendered it invulnerable to the SIV's Nef-based counteraction [21], and in HIV-1 groups M and N, the Vpu protein adapted to counteract Tetherin through its transmembrane domain [22]. In gorillas, the Nef protein adapted to counteract its host's Tetherin but also maintained its ability to counteract the chimpanzee Tetherin [42,43].

Our results agree with the described anti-Tetherin adaptation events: a high proportion of group M rate decelerations were found in the Vpu protein of group M, most of them in the identified Tetherin-binding domain (Table 1 and Figure 3). In addition, in HIV-1 group M the Nef protein was identified with three rate accelerations, which could be interpreted as relaxation of the SIV-Nef anti-Tetherin activity. Numerous group O Nef rate acceleration events were found as well, possibly indicating the loss of chimpanzee/gorilla anti-Tetherin function. A preponderance of acceleration and deceleration events in Nef were also found in the lineages leading to the ancestor of SIVgor/group O, and to SIVcpz*pts*. Thus, our analysis pinpoints sites that may be responsible for the crucial changes in function of Vpu and Nef throughout SIVcpz and HIV evolution (Table S3).
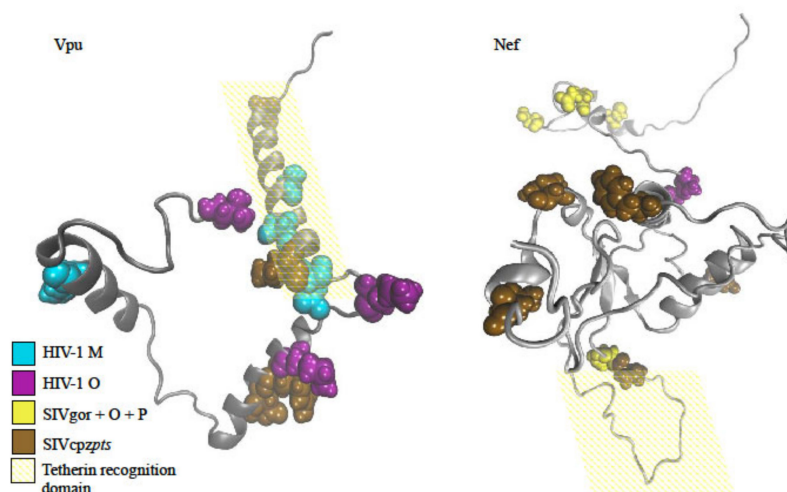

Figure 3. Projection of the identified rate decelerating sites for the prominent branches onto the structure of Vpu (**left**, PDB ID: 2N28) and Nef (**right**, obtained from [44]).

We move on to examine APOBEC3G (A3G), a broad-range antiviral protein that is counteracted by the Vif proteins of Lentiviruses [45,46]. A3G variants of human, chimpanzee, and gorilla have been studied, and it has been found that the Vif recognition domain is conserved between human and chimpanzee A3G proteins, but slightly differs in gorillas [47]. Indeed, the SIVgor adaptation event to

the different host A3G has been demonstrated [48,49]. In line with these results, our analysis shows no enrichment of rate decelerations in M-Vif, yet an excess of rate decelerations in the lineage of HIV-1 group O + SIVgor (Tables S2 and S3).

O-Rev protein stood out for an excess of rate decelerations in the branch leading to group O (Table 1). This is intriguing, since Rev is not known to be involved in any antiviral-related host activity. Some but not all of the decelerations identified are in genomic loci where *rev* rate decelerations overlap with *env* rate decelerations, and this might explain some of this pattern. It remains to be further investigated why the O-Rev protein experienced so many rate decelerations.

3.5. Discussion

Our results tentatively indicate that accessory proteins may have an important role in enabling adaptation to new species in the case of HIV [12]. This suggests that a major common barrier for a host species jump in HIV/SIV is the stage where the virus must overcome intracellular host defenses. Our results further support a lineage-specific model of human adaptation: groups M and O underwent prominent rate shifts in different proteins. This model is supported by the fact that activity of group M and O proteins are indeed different (e.g., Vpu) [42,43], and by the fact that each group originated from an SIV from a different primate.

We were surprised to note that group O had almost twice the amount of rate-shift events as compared to any other lineage analyzed (Table 1). Notably, group O is a non-pandemic strain that remained localized mainly to infections in west-central Africa [50,51]. One possible explanation is that group O entered the human population later than group M, and accordingly, more sites in group O appear as conserved, since they had less time to evolve. We consider this explanation as unlikely, since (a) our phylogenetic model takes into account branch lengths and (b) in such a case, we would also expect less rate accelerations in group O than in group M, whereas we find a very large number of rate accelerations as well. Another possible interpretation may be the relative height of the species barrier a strain had to overcome, as chimpanzees from which group M originated are genetically closer to human than gorillas from which group O originated [52,53]. A species barrier may potentially induce more entrenchment and may also lead to adaptation, as has been described previously [5,54].

3.6. Limitations

This study has several limitations. First, the availability of fully sequenced SIVcpz, SIVgor, and non-HIV-M genomes is still low, reducing the statistical power of the analysis. Indeed, we noted that increasing the sample size of group M sequences that are much more available led to the detection of more rate-shifting sites. We expect that the availability of additional SIVgor sequences will increase the number of rate-shift events that are unique in that clade, possibly revealing more rate changing sites for this lineage. Second, the method that we utilized to identify rate shifts is calibrated to identify major changes in the evolutionary rates along a lineage. Accordingly, it cannot detect sites where a "content shift" occurred, i.e., the rates of substitution remained low in two complementary clades, but the amino acid itself has changed. This is partially demonstrated in Gag_{30}, a previously found marker of HIV adaptation where the rate of evolution changed mildly, while the amino acid in this site changed between chimpanzee and human viruses (Figure S1) [18]. Accordingly, we have detected this site but were inconclusive regarding the lineage where the rate shift occurred, and thus, it was not reported in this study. In addition, our method could not detect adaptation by amino acid insertion or deletion events, and we did not analyze for rate shifts at the RNA level and thus, for example, did not detect rate shifts at RNA structures of the virus.

4. Conclusions

We used a phylogenetic modeling approach to perform a large scale analysis of the constraints operating in all proteins of the HIV genome, across different lineages of HIV-1 and SIV. We inferred individual amino acids where rate shifts occurred and characterized the distribution of rate shifts across

different proteins and across the tree. We found that, in many cases, the number of rate accelerations and rate decelerations is similar and suggest that epistatic interactions are pervasive in the evolution of HIV-1 and SIV proteins, and that these likely led to many of the inferred rate shifts observed herein. Nevertheless, our results do suggest that at least some of the rate shifts that we detected may be due to selection and may be related to the cross-species transmission of the virus from apes into humans. Our results also implicate several of the accessory proteins of HIV as mediators that allow cross-species transmission events to occur. Our compiled list of the positions suggested as rate shifting can now serve as a basis for future functional studies that will allow a better and deeper understanding of the constraints operating on HIV-1 and SIV proteins.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4915/12/11/1312/s1. Figure S1: Residues distribution for Gag30 along the HIV-1/SIV phylogeny. This position exhibits a pattern more similar to "content shift" than to "rate shift", Table S1: Maximum log-likelihood (LL) values for the analysis of the nine HIV-1/SIVcpz/SIVgor proteins under the rate shift and null models, Table S2: Rate shifts as percentage from total rate shifts for prominent branches, for rate decelerating sites (upper) and rate accelerating sites (lower), colored by intensity. Data presented in this table is not normalized by protein length, Table S3: Rate deceleration events that may be related to previously reported species-specific adaptation events, File S1: Phylogenetic tree and sequence alignments used for identifying rate shifting sites. The zip file contains all nine alignments and constructed phylogeny used for identifying rate shifting sites, File S2: HIV/SIV sites identified as rate shifting. Positions are provided in HXB2 reference sequence coordinates. Branch number field refer to the branch number outputted by RASER (Penn et al. 2008); for ease of reading we provide branch labels for most branches, File S3: HIV/SIV sites identified as rate shifting when group M contains many more sequenced than group O. Positions are provided in HXB2 reference sequence coordinates. Branch number field refer to the branch number outputted by RASER (Penn et al. 2008); for ease of reading we provide branch labels for most branches.

References

1. Hahn, B.H.; Shaw, G.M.; De Cock, K.M.; Sharp, P.M. AIDS—AIDS as a zoonosis: Scientific and public health implications. *Science* **2000**, *287*, 607–614. [CrossRef] [PubMed]

2. Van Heuverswyn, F.; Li, Y.; Neel, C.; Bailes, E.; Keele, B.F.; Liu, W.; Loul, S.; Butel, C.; Liegeois, F.; Bienvenue, Y.; et al. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **2006**, *444*, 164. [CrossRef] [PubMed]

3. D'Arc, M.; Ayouba, A.; Esteban, A.; Learn, G.H.; Boue, V.; Liegeois, F.; Etienne, L.; Tagg, N.; Leendertz, F.H.; Boesch, C.; et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1343–E1352. [CrossRef] [PubMed]

4. Takehisa, J.; Kraus, M.H.; Ayouba, A.; Bailes, E.; Van Heuverswyn, F.; Decker, J.M.; Li, Y.; Rudicell, R.S.; Learn, G.H.; Neel, C.; et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J. Virol.* **2009**, *83*, 1635–1648. [CrossRef]

5. Bell, S.M.; Bedford, T. Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission. *PLoS Pathog.* **2017**, *13*, e1006466. [CrossRef]

6. Korber, B.; Muldoon, M.; Theiler, J.; Gao, F.; Gupta, R.; Lapedes, A.; Hahn, B.H.; Wolinsky, S.; Bhattacharya, T. Timing the ancestor of the HIV-1 pandemic strains. *Science* **2000**, *288*, 1789–1796. [CrossRef]

7. Lemey, P.; Pybus, O.G.; Rambaut, A.; Drummond, A.J.; Robertson, D.L.; Roques, P.; Worobey, M.; Vandamme, A.M. The molecular population genetics of HIV-1 group O. *Genetics* **2004**, *167*, 1059–1068. [CrossRef]

8.   Wertheim, J.O.; Worobey, M. Dating the Age of the SIV Lineages That Gave Rise to HIV-1 and HIV-2. *PLoS Comp. Biol.* **2009**, *5*, e1000377. [CrossRef]

9.   Bailes, E.; Gao, F.; Bibollet-Ruche, F.; Courgnaud, V.; Peeters, M.; Marx, P.A.; Hahn, B.H.; Sharp, P.M. Hybrid origin of SIV in chimpanzees. *Science* **2003**, *300*, 1713. [CrossRef]

10.   Leitner, T.; Dazza, M.C.; Ekwalanga, M.; Apetrei, C.; Saragosti, S. Sequence Diversity among Chimpanzee Simian Immunodeficiency Viruses (SIVcpz) Suggests that SIVcpz Pts Was Derived from SIVcpz Ptt through Additional Recombination Events. *AIDS Res. Hum. Retrovir.* **2007**, *23*, 1114–1118. [CrossRef]

11.   Li, Y.; Ndjango, J.B.; Learn, G.H.; Ramirez, M.A.; Keele, B.F.; Bibollet-Ruche, F.; Liu, W.; Easlick, J.L.; Decker, J.M.; Rudicell, R.S.; et al. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *J. Virol.* **2012**, *86*, 10776–10791. [CrossRef] [PubMed]

12.   Etienne, L.; Hahn, B.H.; Sharp, P.M.; Matsen, F.A.; Emerman, M. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* **2013**, *14*, 85–92. [CrossRef] [PubMed]

13.   Longdon, B.; Brockhurst, M.A.; Russell, C.A.; Welch, J.J.; Jiggins, F.M. The evolution and genetics of virus host shifts. *PLoS Pathog.* **2014**, *10*, e1004395. [CrossRef] [PubMed]

14.   Saéz, A.M.; Weiss, S.; Nowak, K.; Lapeyre, V.; Zimmermann, F.; Düx, A.; Kühl, H.S.; Kaba, M.; Regnaut, S.; Merkel, K.; et al. Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Mol. Med.* **2015**, *7*, 17–23. [CrossRef] [PubMed]

15.   Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef] [PubMed]

16.   Yuan, Z.; Kang, G.; Ma, F.; Lu, W.; Fan, W.; Fennessey, C.M.; Keele, B.F.; Li, Q. Recapitulating cross-species transmission of simian immunodeficiency virus SIVcpz to humans by using humanized BLT mice. *J. Virol.* **2016**, *90*, 7728–7739. [CrossRef]

17.   Sato, K.; Misawa, N.; Takeuchi, J.S.; Kobayashi, T.; Izumi, T.; Aso, H.; Nagaoka, S.; Yamamoto, K.; Kimura, I.; Konno, Y.; et al. Experimental Adaptive Evolution of Simian Immunodeficiency Virus SIVcpz to Pandemic Human Immunodeficiency Virus Type 1 by Using a Humanized Mouse Model. *J. Virol.* **2018**, *92*, e01905–e01917. [CrossRef]

18.   Wain, L.V.; Bailes, E.; Bibollet-Ruche, F.; Decker, J.M.; Keele, B.F.; Van Heuverswyn, F.; Li, Y.; Takehisa, J.; Ngole, E.M.; Shaw, G.M.; et al. Adaptation of HIV-1 to its human host. *Mol. Biol. Evol.* **2007**, *24*, 1853–1860. [CrossRef]

19.   Yamada, E.; Nakaoka, S.; Klein, L.; Reith, E.; Langer, S.; Hopfensperger, K.; Iwami, S.; Schreiber, G.; Kirchhoff, F.; Koyanagi, Y.; et al. Human-Specific Adaptations in Vpu Conferring Anti-tetherin Activity Are Critical for Efficient Early HIV-1 Replication In Vivo. *Cell Host Microbe* **2018**, *23*, 110.e117–120.e117. [CrossRef]

20.   Neil, S.J.D.; Zang, T.; Bieniasz, P.D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **2008**, *451*, 425–430. [CrossRef]

21.   Zhang, F.; Wilson, S.J.; Landford, W.C.; Virgen, B.; Gregory, D.; Johnson, M.C.; Munch, J.; Kirchhoff, F.; Bieniasz, P.D.; Hatziioannou, T. Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell Host Microbe* **2009**, *6*, 54–67. [CrossRef] [PubMed]

22.   Sauter, D.; Schindler, M.; Specht, A.; Landford, W.N.; Munch, J.; Kim, K.A.; Votteler, J.; Schubert, U.; Bibollet-Ruche, F.; Keele, B.F.; et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* **2009**, *6*, 409–421. [CrossRef] [PubMed]

23.   Van Damme, N.; Goff, D.; Katsura, C.; Jorgenson, R.L.; Mitchell, R.; Johnson, M.C.; Stephens, E.B.; Guatelli, J. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* **2008**, *3*, 245–252. [CrossRef] [PubMed]

24.   Knudsen, B.; Miyamoto, M.M. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14512–14517. [CrossRef] [PubMed]

25.   Penn, O.; Stern, A.; Rubinstein, N.D.; Dutheil, J.; Bacharach, E.; Galtier, N.; Pupko, T. Evolutionary modeling of rate shifts reveals specificity determinants in HIV-1 subtypes. *PLoS Comput. Biol.* **2008**, *4*, e1000214. [CrossRef] [PubMed]

26.   Galtier, N. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* **2001**, *18*, 866–873. [CrossRef] [PubMed]

27. Lopez, P.; Casane, D.; Philippe, H. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* **2002**, *19*, 1–7. [CrossRef]

28. Breen, M.S.; Kemena, C.; Vlasov, P.K.; Notredame, C.; Kondrashov, F.A. Epistasis as the primary factor in molecular evolution. *Nature* **2012**, *490*, 535–538. [CrossRef]

29. Starr, T.N.; Thornton, J.W. Epistasis in protein evolution. *Protein Sci.* **2016**, *25*, 1204–1218. [CrossRef]

30. Bridgham, J.T.; Ortlund, E.A.; Thornton, J.W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **2009**, *461*, 515–519. [CrossRef]

31. Pollock, D.D.; Thiltgen, G.; Goldstein, R.A. Amino acid coevolution induces an evolutionary Stokes shift. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1352–E1359. [CrossRef] [PubMed]

32. Shah, P.; McCandlish, D.M.; Plotkin, J.B. Contingency and entrenchment in protein evolution under purifying selection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3226–E3235. [CrossRef] [PubMed]

33. Wu, N.C.; Thompson, A.J.; Xie, J.; Lin, C.W.; Nycholat, C.M.; Zhu, X.; Lerner, R.A.; Paulson, J.C.; Wilson, I.A. A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. *Nat. Commun.* **2018**, *9*, 1–13. [CrossRef]

34. Da Silva, J.; Coetzer, M.; Nedellec, R.; Pastore, C.; Mosier, D.E. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* **2010**, *185*, 293–303. [CrossRef] [PubMed]

35. Haddox, H.K.; Dingens, A.S.; Hilton, S.K.; Overbaugh, J.; Bloom, J.D. Mapping mutational effects along the evolutionary landscape of HIV envelope. *Elife* **2018**, *7*, e34420. [CrossRef] [PubMed]

36. Kuiken, C.; Foley, B.; Leitner, T.; Apetrei, C.; Hahn, B.; Mizrachi, I.; Mullins, J.; Rambaut, A.; Wolinsky, S.; Korber, B. *HIV Sequence Compendium 2010*; Los Alamos National Laboratory (LANL): Los Alamos, NM, USA, 2010.

37. Loytynoja, A.; Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10557–10562. [CrossRef]

38. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef] [PubMed]

39. Schultz, A.K.; Zhang, M.; Bulla, I.; Leitner, T.; Korber, B.; Morgenstern, B.; Stanke, M. jpHMM: Improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* **2009**, *37*, W647–W651. [CrossRef]

40. McNatt, M.W.; Zang, T.; Hatziioannou, T.; Bartlett, M.; Fofana, I.B.; Johnson, W.E.; Neil, S.J.; Bieniasz, P.D. Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog.* **2009**, *5*, e1000300. [CrossRef]

41. Evans, D.T.; Serra-Moreno, R.; Singh, R.K.; Guatelli, J.C. BST-2/tetherin: A new component of the innate immune response to enveloped viruses. *Trends Microbiol.* **2010**, *18*, 388–396. [CrossRef]

42. Lim, E.S.; Malik, H.S.; Emerman, M. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J. Virol.* **2010**, *84*, 7124–7134. [CrossRef] [PubMed]

43. Yang, S.J.; Lopez, L.A.; Exline, C.M.; Haworth, K.G.; Cannon, P.M. Lack of adaptation to human tetherin in HIV-1 group O and P. *Retrovirology* **2011**, *8*, 78. [CrossRef] [PubMed]

44. Geyer, M.; Peterlin, B.M. Domain assembly, surface accessibility and sequence conservation in full length HIV-1 Nef. *FEBS Lett.* **2001**, *496*, 91–95. [CrossRef]

45. Mangeat, B.; Turelli, P.; Caron, G.; Friedli, M.; Perrin, L.; Trono, D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **2003**, *424*, 99–103. [CrossRef] [PubMed]

46. Sheehy, A.M.; Gaddis, N.C.; Malim, M.H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **2003**, *9*, 1404. [CrossRef] [PubMed]

47. Russell, R.A.; Smith, J.; Barr, R.; Bhattacharyya, D.; Pathak, V.K. Distinct Domains within APOBEC3G and APOBEC3F Interact with Separate Regions of Human Immunodeficiency Virus Type 1 Vif. *J. Virol.* **2009**, *83*, 1992–2003. [CrossRef]

48. Compton, A.A.; Emerman, M. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathog.* **2013**, *9*, e1003135. [CrossRef]

49. Letko, M.; Silvestri, G.; Hahn, B.H.; Bibollet-Ruche, F.; Gokcumen, O.; Simon, V.; Ooms, M. Vif Proteins from Diverse Primate Lentiviral Lineages Use the Same Binding Site in APOBEC3G. *J. Virol.* **2013**, *87*, 11861–11871. [CrossRef]

50. Peeters, M.; Gueye, A.; Mboup, S.; Bibollet-Ruche, F.; Ekaza, E.; Mulanga, C.; Ouedrago, R.; Gandji, R.; Mpele, P.; Dibanga, G.; et al. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **1997**, *11*, 493–498. [CrossRef]

51. Leoz, M.; Feyertag, F.; Kfutwah, A.; Mauclere, P.; Lachenal, G.; Damond, F.; De Oliveira, F.; Lemee, V.; Simon, F.; Robertson, D.L.; et al. The Two-Phase Emergence of Non Pandemic HIV-1 Group O in Cameroon. *PLoS Pathog.* **2015**, *11*, e1005029. [CrossRef]

52. Hasegawa, M.; Kishino, H.; Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **1985**, *22*, 160–174. [CrossRef] [PubMed]

53. Ruvolo, M. Molecular phylogeny of the hominoids: Inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* **1997**, *14*, 248–265. [CrossRef] [PubMed]

54. Charleston, M.A.; Robertson, D.L. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst. Biol.* **2002**, *51*, 528–535. [CrossRef] [PubMed]