bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evolutionary rate shifts suggest species-specific adaptation events in HIV-1 and SIV

Maoz Gelbart and Adi Stern

School of Molecular Cell Biology and Biotechnology, George S. Wise faculty of Life Sciences, Tel Aviv University, Israel

* Correspondence to AS: sternadi@post.tau.ac.il

Abstract

The process of molecular adaptation following a cross-species virus transmission event is currently poorly understood. Here, we identified 137 protein sites that experienced deceleration in their rate of evolution along the HIV-1/SIV phylogeny, likely indicating gain-of-function and consequent adaptation. The majority of such events occurred in parallel to cross-species transmission events and varied between HIV-1 groups, indicating independent adaptation strategies. The evolutionary rate decelerations we found were particularly prominent in accessory proteins that counteract host antiviral restriction factors, suggesting that these factors are a major barrier to viral adaptation to a new host. Surprisingly, we observed that the non-pandemic HIV-1 group O, derived from gorillas, exhibited more rate deceleration events than the pandemic group M, derived from chimpanzees. We suggest that the species barrier is higher when the genetic distance of the hosts increases. Our approach paves the way for subsequent studies on cross-species transfers in other major pathogens.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 Virus (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 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 inception of this virus in the western lowland gorilla population somewhere in the 19th century [4]. SIVcpz itself was found to be a transmission from other primates, dated at roughly 1500 and leading to the 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 species barrier is a subject of much interest, since many pandemic human viruses arose from zoonosis events such as the influenza strain of "Spanish flu" H1N1, measles virus (MeV), and SARS coronavirus [13]. Due to genetic differences between the hosts, virus adaptation occurs at multiple levels: at the level of entry to target cells; interaction with the host adaptive immune system; interaction with host antiviral restriction factors; and recruitment of host cellular machinery by the virus [14-16]. Protein adaptations are reflected in the history of genomes and may be manifested in changes in the amino acid substitution rates of the adaptive sites, which are expected to be more conserved in the clade where adaptation

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45 happened [17]. This is due to new roles gained by these protein sites that constrain their
46 evolution. A reciprocal phenomenon where amino acids are less conserved in one clade than in
47 other clades is also possible, and may either reflect a loss of function or a gain of function
48 manifested as positive diversifying selection. Identification of such "rate shifting sites" can thus
49 reveal virus adaptation events, and is expected to promote a better understanding of cross-
50 species transmissions in general and in the HIV pandemic in particular in this study.

51 A method for identification of evolutionary rate changes has been previously established and
52 was used to study intra-subtype HIV adaptation events [18]. To date, the lack of SIVcpz and
53 group O full genomic sequences limited the ability to study adaptation across cross-species
54 transmission events. Here, we utilized the growing availability of diverse and full HIV-1 and SIV
55 genomes to identify many sites in various SIVcpz/SIVgor/HIV-1 clades whose evolutionary rate
56 changed across clades [19]. We demonstrate how cross-species transmissions are correlated
57 with abundant evolutionary rate shifts and how known adaptation events are manifested in
58 different amino acid substitution rates between lineages. Based on the rate shift patterns, we
59 suggest previously unknown adaptation events, and highlight the exceptional amount of
60 evolutionary rate shifts observed in HIV-1 group O, possibly due to a more extreme host species
61 barrier.

62 Results

63 We have previously developed the RASER tool to identify sites that display change in the rate of
64 evolution along a given branch of a phylogenetic tree (see example in Figure 1) [18]. In essence,
65 this tool takes a phylogeny and a multiple sequence alignment and contrasts the rates of
66 evolution of amino acids along all branches in the phylogeny, to determine if the evolutionary
67 rates of some amino acids in some branches are better explained by a model that allows for

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68 evolutionary rate changes. We queried the Los Alamos HIV database [19] for all available HIV-1,
69 SIVcpz and SIVgor full sequences, from which we took the sequences of all nine HIV-1 proteins.
70 Due to high representation of group M sequences, we downsampled this group to the size of
71 available group O sequences, retaining the strains with the most internal variation (total number
72 of sequences N=126, see Methods). A different downsample of group M sequences to a bigger
73 size was also conducted, to validate results robustness to cohort size variation (N=223, see
74 Methods). All site coordinates are reported based on the HIV-1 subtype B reference strain HXB2
75 (Methods).

76 **Most rate shift events were identified in speciation branches.** We first sought to characterize
77 the branches where rate shift events were suggested. A total of 271 rate shifting positions were
78 identified along the phylogeny. 137 of them were identified as rate-deceleration events and an
79 additional 134 were identified as rate-accelerations. As listed in Table 1, 230 out of the 271 rate
80 shifting positions were attributed to four branches: the branch separating group M from SIVcpz,
81 the branch separating group O from SIVgor, the branch separating SIVgor and group O from
82 SIVcpz and the branch separating SIVcpz*pts* from other viruses. Likelihood ratio tests were found
83 to be highly significant in favor of a rate shift model across all HIV proteins (Supplementary
84 Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

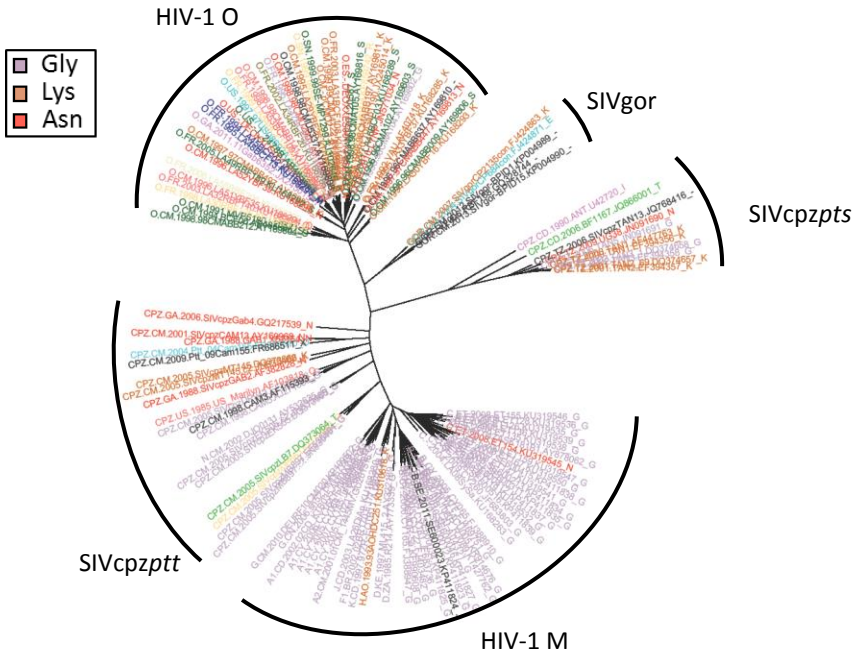


Figure 1. Projection of Env_{458} on the HIV-1/SIV phylogeny. Each leaf (corresponding to an HIV/SIV strain) is color coded based on the amino acid present at Env_{458}. The evolutionary rate of Env_{458} was found to be slower in HIV-1 group M than in the rest of the phylogeny.

Different rate deceleration patterns identified in different branches. A special focus was given to rate deceleration events, as they likely indicate a gain of function in a protein. In Figure 2 we mapped the percent of rate shifts in each protein observed in each lineage, corresponding to different virus speciation events. While rate deceleration distribution patterns differed between the different lineages (see Figure 2 and Supplementary Table 1), one common theme was that all prominent branches underwent a significant portion of their rate deceleration events in the Env protein. The next two common rate decelerating proteins were Gag and Pol identified in all four highlighted branches, which is expected given that together with Env these are the longest proteins in HIV/SIV. When controlling for protein size (as in Table 1), a different pattern emerged: Vpu seems to have undergone a disproportionate portion of rate deceleration events, in particular in the branch leading to group M. In general, the rate deceleration patterns for the non-structural proteins seemed to be branch-specific: In the branch leading to SIVgor and group

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

98 O, Nef and Vif experienced the most relative rate deceleration events. In the branch leading to
99 group M, Vpu was responsible for the most rate deceleration events; and in the branch leading
100 to group O, Rev contributed a significant portion of all rate decelerations for that branch
101 (Supplementary Table 1). The branch separating SIVcpz*pts* from other viruses displayed a more
102 diverse pattern with relatively less rate decelerations in the Env protein and more rate
103 decelerations in the Nef, Vif and Vpu proteins.

104 **Table 1. Rate shifts as *percentage from total protein size* for prominent branches, for rate**
105 **decelerating sites (upper) and rate accelerating sites (lower), colored by intensity.** The raw
106 data underlying this table are provided in Supplementary File 2.

DEC	Structural proteins			Regulatory proteins		Accessory proteins				
Branch/protein	**Env**	**Gag**	**Pol**	**Rev**	**Tat**	**Nef**	**Vif**	**Vpr**	**Vpu**	**Total number of mutations**
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 mutations**
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

107

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

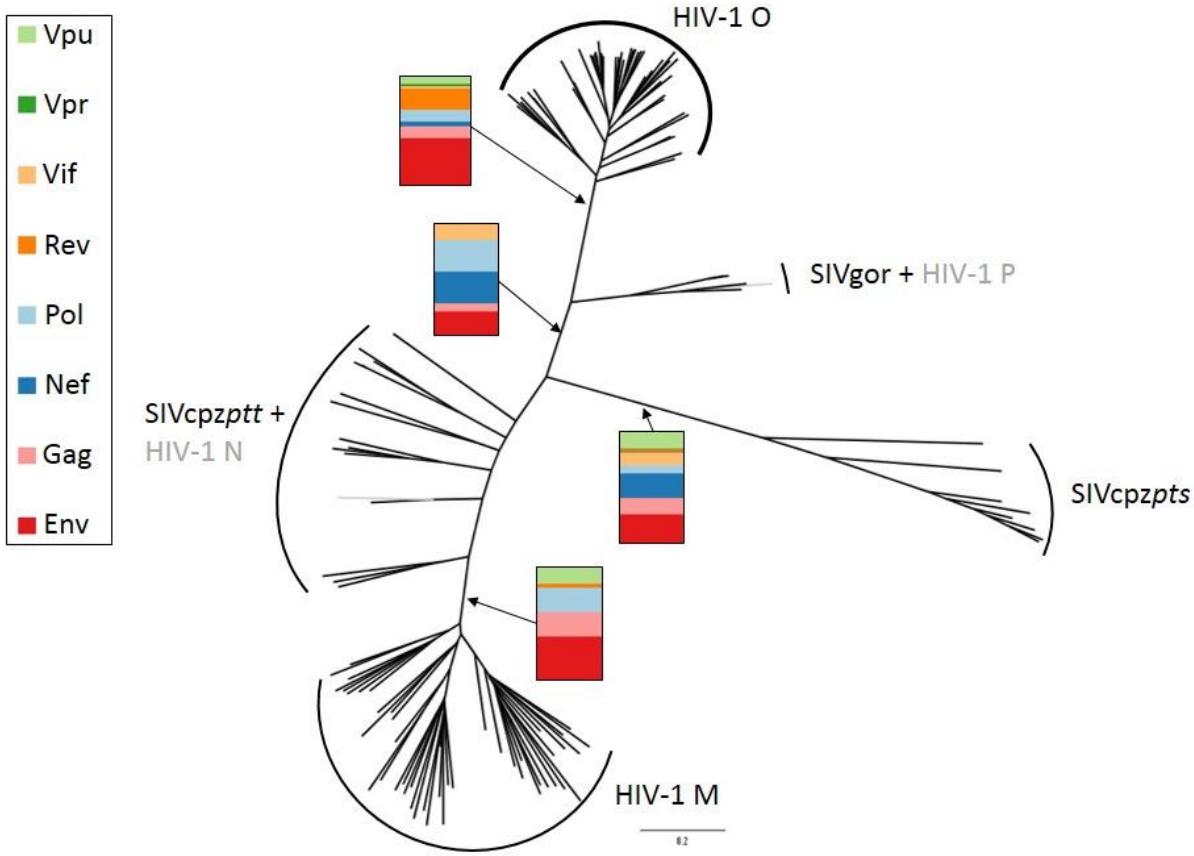


Figure 2. Proposed rate deceleration patterns in prominent branches, shown on the phylogeny. Rate decelerations are shown as percentage from total rate decelerations for each branch.

Parallel rate shift events in HIV-1 groups M and O. Rate shift events common to several HIV-1 groups are of special interest, since they may better reflect the species barrier in general and not lineage-specific adaptation event. As is evident from Table 2, most parallel rate shift events of HIV-1 groups O and M (or their closest precursor branch) were rate deceleration events, half of which occurred in the highly conserved Pol protein. Notably, the consensus amino acid identity in those parallel-occurring rate deceleration events varied between the M and O clade for all identified positions, suggesting once again that different adaptation strategies occurred in each lineage. Only a single parallel rate acceleration event was identified, in Nef_8 of HIV-1

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

120 groups M and O. This may indicate a relaxation of prior constraint as the SIV allele was
121 maintained in many HIV-1 sequences.

122 **Table 2. Rate shifting sites identified independently in parallel in HIV-1 groups M and O or**
123 **their immediate ancestor.**

Protein	Rate Shifting site	Rate shift type	Branches	Consensus Amino Acids
Pol	349	Deceleration	M, Pre O	M: E, Pre O: P
Pol	406	Deceleration	Pre M, O	Pre M: S, O: V
Pol	926	Deceleration	M, Pre O	M: K, Pre O: T
Env	61	Deceleration	M, O	M: Y, O: T
Env	772	Deceleration	M, O	M: R, O: S
Rev	71	Deceleration	Pre M, O	Pre M: V, O: N
Nef	8	Acceleration	M, O	NA

124 NA- Not applicable

125 **Rate shift events are often parallel to known macromolecular adaptation events.** Intense
126 research into the origins of the HIV-1 pandemic and its sources identified several major
127 adaptation events of HIV-1 to its human host. These events include adapting to cellular
128 variations such as differences in host receptor CD4; overcoming host innate immune restriction
129 factors such as Tetherin (BST-2) and APOBEC3G; and of course, evasion from the adaptive
130 immune system [16, 20-25]. We therefore sought to evaluate whether the proposed rate shifts
131 agree with the known adaptation events.

132 **Anti Tetherin/BST-2 adaptations.** Tetherin/BST-2 is a trans-membrane antiviral
133 restriction factor that disrupts the budding phase of the retroviral lifecycle, thus preventing the
134 infection from spreading [26, 27]. SIVcpz and many more SIV's antagonize Tetherin through
135 interactions of their nef protein with the cytoplasmic tail of Tetherin [28, 29]. Tetherin is a fast
136 evolving protein, and differs between chimpanzee, gorilla and human [30, 31]. In humans, a
137 significant deletion in the cytoplasmic tail of Tetherin rendered it invulnerable to the SIV's Nef-
138 based counteraction [29]. Therefore, in order to regain infectivity HIV-1 developed an

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alternative way to antagonize Tetherin. In groups M and N, Vpu protein adapted to antagonize Tetherin through its transmembrane domain [32]. In group O the virus adapted its Nef protein to counteract the human Tetherin, albeit with less efficiency than in group M [33, 34]. Notably, a single O strain that uses its Vpu protein to encounter human Tetherin has been described [35]. No anti-Tetherin adaptation event is known to have occurred in the rare group P [36]. In gorilla, the Nef protein adapted to encounter its host's Tetherin but also maintained its ability to encounter the chimpanzee tetherin [37, 38].

Our results support the described anti-Tetherin adaptation events: a dramatically high proportion (14% of all identified group M rate decelerations) were found in the Vpu protein, most of them in the Tetherin-binding domain (Supplementary Table 1 and Figure 3). In addition, in HIV-1 group M the Nef protein was identified with three rate accelerations, supporting the loss of function of its anti-Tetherin activity. This is also supported by a reintroduction experiment of HIV-1 group M strain into chimpanzee in which M-Nef readapted to chimpanzee host, as the mutations required for M-Nef to regain anti-Tetherin activity in chimpanzees (Nef_{163} and Nef_{169}) [39] were located proximally to Nef_{157}, one of the rate accelerating sites of M-Nef identified in our study. Numerous O-Nef rate acceleration events were found as well, supporting the loss of chimpanzee/gorilla anti-Tetherin function. A preponderance of acceleration and deceleration events were also found in the lineages leading to the ancestor of SIVgor and O and to SIVcpz*pts*; one of those rate deceleration events is at Nef_{177}, which is proximal to the C-loop region which has been shown to be related to O-Nef anti-Tetherin activity [40]. Thus, our analysis pinpoints the sites that are likely responsible for the dramatic changes in function of Vpu and Nef throughout SIVcpz and HIV evolution.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

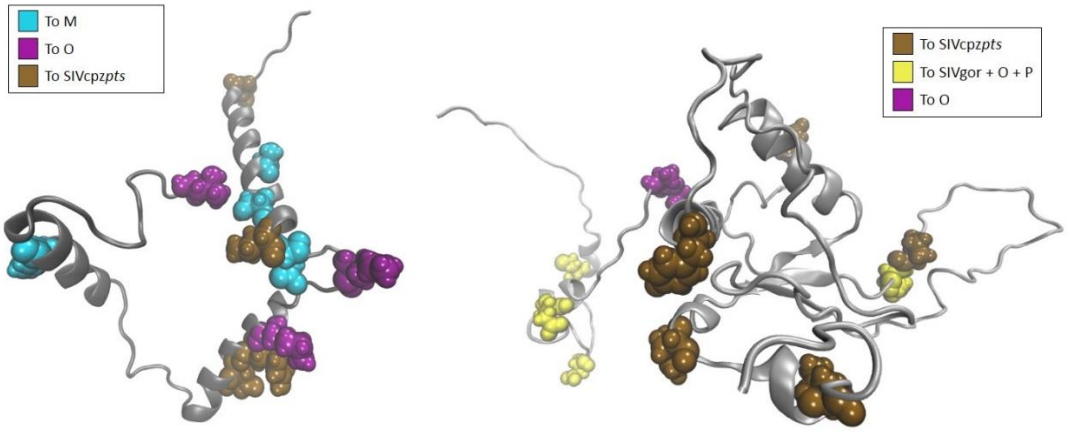


Figure 3. Projection of the identified rate decelerations onto the tertiary structure of Vpu protein (left, PDB ID: 2N28) and Nef protein (right, obtained from [41]) in the major lineages where it was identified.

Anti APOBEC3G adaptations. APOBEC3G (A3G) is a broad-range antiviral protein that is packaged into HIV-1 virions, and upon reverse transcription adds C→U mutations to the synthesized DNA strand, thus potentially creating nonviable genomes [42]. Lentiviruses use their Vif protein to overcome this factor by degrading it as well as other antiviral proteins from the APOBEC3 family (such as A3F, A3D) [43]. A3G variants of human, chimpanzee and gorilla have been studied and the Vif recognition domain has been identified at A3G residues 126-132 [44]. Of those species, gorilla has a different residue at A3G position 129 as compared to that of human and chimpanzee A3Gs and indeed the SIVgor adaptation event to the different host A3G has been demonstrated by [45, 46]. It has been shown that SIVgor can replicate in the presence of human-, gorilla- and chimpanzee-A3G, but HIV-1 group M or SIVcpz cannot replicate in the presence of gorilla-A3G, indicative of a gain-of-function in the adaptation process of SIVgor to gorilla [46].

In line with these findings, our analysis shows that the branch leading to SIVgor and HIV-1 groups O and P experienced 14% of its rate decelerations in the Vif protein, particularly in

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

179 positions Vif_{73} and Vif_{167}. In HIV-1, Vif_{73} is in close proximity to the A3G-binding motif Vif_{69-72} that
180 is also responsible in-part for interaction with APOBEC3F (which in itself contains differences in
181 the Vif interacting loop between chimpanzee and human/gorilla) [47-49]. O-Vif experienced a
182 single rate deceleration in position Vif_{127}, which is located in the Cullin-5 interacting domain [48].
183 Lack of rate deceleration events in M-Vif suggest no adaptation of this protein in this strain to
184 the human host; this is supported by the observation that SIVcpz-Vif can encounter human A3G
185 [46].

186 **Env adaptations.** Another major barrier for host jumps lies at the cell entry level. CD4 is
187 the target receptor of SIV and HIV, while CCR5 or CXCR4 act as co-receptors, recognized by the
188 viral Env protein which is exposed to the virus external surface. Previous studies mapped the
189 genetic diversity between different chimpanzees' CD4 and revealed that there are several
190 differences between chimpanzee subspecies as well as differences between chimpanzee versus
191 gorilla and human, especially in the regions that are in close contact with Env subunit gp120 [50,
192 51]. Some of those changes affect the glycosylation patterns of CD4, making these differences
193 even more distinguishable [50, 51]. It is therefore likely that HIV-1 had to adapt to recognize
194 these differences in CD4.

195 Our analysis overall supports a high level of deceleration and acceleration events found
196 in Env, possibly reflecting the changes that occurred in the host CD4. For instance, many of the
197 changes in M-Env occurred in CD4 interacting residues of Env_{374}, Env_{455}, Env_{456} and Env_{458} [52].
198 However, we cannot rule out that the very high rate of evolution in Env, driven by the adaptive
199 immune system [53], have led to changes that “mimic” a rate shift. Indeed when accounting for
200 the size of the Env protein, it seems that this protein is less pronounced in its rate shift
201 distribution (Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gag_{30} human adaptation marker. Position 30 of Gag polyprotein has been previously suggested to be an adaptation event of HIV-1 to its human host, as it diversified from Methionine in SIVcpz*ptt* and SIVgor strains into Lysine or Arginine in many HIV-1 strains [54]. Reintroduction of HIV-1 to chimpanzees resulted in reversion of this position back to the conserved Met of SIVcpz*ptt* and therefore it was suggested as an adaptation event to the human host [54]. Notably, subtype C of group M is conserved for Methionine at Gag_{30}, suggesting that adaptation occurred only in some subtypes.

Our results indeed suggest an evolutionary rate shift event in Gag_{30} yet are not conclusive about the branches in which the rate shift happened. This is likely since this site displays a "content shift" rather than a strong "rate shift" (elaborated in the discussion).

Table 3. Summary of HIV rate deceleration events related to known antiviral activity.

Viral Lineage	Viral Protein	Rate decelerating positions identified	Rate deceleration(s) related to the following restriction factor	Function of the restriction factor
Group M	Vpu	18, 23, 28, 66	Tetherin	Prevents virus release from infected cells
SIVgor and Group O	Nef	20, 28, 38, 49, 50, 177	Tetherin	Prevents virus release from infected cells
SIVgor and group O	Vif	73, 127, 167	APOBEC3G	Induces hypermutation of viral genomes

Discussion

Understanding the molecular changes in a pathogen when adapting to infect a new host species is of high importance. To the best of our knowledge, our study represents the first large-scale approach to detect adaptation events in the transition from non-human primates to humans, and relies on a robust phylogenetic modeling approach. We discovered many lineage-specific adaptation-like events in many proteins of HIV-1, SIVgor and SIVcpz*pts*, and have reported sites in where groups M and O suggested to undergone parallel rate shifts. The majority of sites

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

220 undergoing rate deceleration were found in the Env protein; when correcting for protein size,
221 the relatively highest number of rate decelerating were found in proteins that adapted to host
222 restriction factors. Accordingly, we suggest that the major common barrier for a host species
223 jump is composed of both the entry stage and the stage where the virus must overcome the first
224 barrier of cellular defenses. Our results further support a differential model of adaptation:
225 groups M and O underwent adaptation in different proteins, and when adaptation occurred at
226 similar sites, the amino-acid was different in both groups. This model is supported by the fact
227 that activity of group M and O proteins are indeed different (e.g., Vpu), and by the fact that each
228 group originated from an SIV from a different primate.

229 Since group M is the pandemic strain of HIV-1, we initially expected that it would experience the
230 highest number of rate shift events among all HIV-1 groups, indicating more efficient adaptation
231 to its human host. Surprisingly, our analysis revealed that group O had almost twice the amount
232 of rate shift events, despite being a non-pandemic strain that remained localized mainly to
233 infections in west-central Africa [55, 56]. We suggest that the number of rate shift events is not
234 the determinant of a pandemic strain; it may rather reflect the relative height of the species
235 barrier a strain had to overcome, as chimpanzee from which group M originated is genetically
236 closer to human than gorilla from which group O originated [57, 58].

237 We further noted a large number of rate shifts that occurred in the lineage separating the two
238 subspecies of SIVcpz (SIVcpz*ptt* and SIVcpz*pts*). Presumably these events correspond to
239 adaptations of the virus to the two chimpanzee subspecies, which are genetically divergent.
240 Further research will be required to understand if and how this has affected the adaptation of
241 the virus to the human host.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Limitations. This study has several limitations. First, the availability of full coverage of SIVcpz and the SIVgor 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 detection of more rate shifting sites (Supplementary File 1). 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 important sites for this lineage. Second, the method that we utilized to identify rate shifts is calibrated to identify dramatic changes in the evolutionary rate along a lineage. Accordingly, it cannot detect sites where a "content shift" occurred, i.e. the amino acid changed and remained conserved in two complementary lineages, since this entails only a minor change in evolutionary rate. This is partially demonstrated in Gag_{30}, where the rate of evolution changed mildly, while the content of this site changed between chimpanzee and human viruses.

Implications. The compiled list of the positions suggested as rate decelerating can now serve as a guide for future functional studies that aim to understand the differences among HIV-1 and SIV proteins. Furthermore, the ability to track adaptation events by utilizing sequence data highlights the power of the method when studying emerging pandemics, strengthening the need to sequence full genomes of pathogens broadly.

Conclusions. Genetic sequences of viruses and specifically HIV-1 and SIV viruses can be harnessed to identify adaptation events of emerging pathogens to their new host species. Our results suggest that innate immunity serves as a strong barrier for cross-species transmission events, and that this barrier imposed a strong selective pressure for viruses to adapt as they crossed these barriers with increasing efficiency.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

In order to collect sequences for this study, the Los Alamos HIV sequence database (available online at http://www.hiv.lanl.gov [19]) 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. Two different datasets were constructed: (i) with a very large number of group M sequences compared to the amount of group O sequences, and (ii) with an equal number of group M and group O sequences. Due to computational reasons, dataset (i) included 200 HIV-1 sequences, most of them group M. In both datasets, the sequences were sampled 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. The results in this study are reported mainly with datasets (ii), chosen since it allows comparing the result from group M and group O. Additional sites found with datasets (i) are reported in Supplementary File 1.

Initial multiple sequence alignments of the nine proteins were performed using PRANK and iteratively improved until convergence [59]. 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 [60]. We next used the reconstructed phylogeny as a guide tree to realign each protein with PRANK. JpHMM was used to validate that the strains used in the analysis are not inter-group recombinants [61].

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to identify evolutionary rate shifts, we used RASER [18] 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 site-specific evolutionary rates. 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 probability higher than 0.6 are considered here as significant. 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.

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, denoted as dataset (i). Analysis revealed more rate decelerations in the branch leading to group M than in dataset (ii) analysis (48 compared to 28, Supplementary File 1). 54% of the positions identified as rate-decelerating in group M in dataset (ii) (n=15) were also identified in dataset (i). 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.

Acknowledgements

The authors would like to thank Dr. Osnat Penn, Prof. Eran Bacharach (Tel Aviv University) and Prof. Tal Pupko (Tel Aviv University) for their contribution to the inception of this study, Prof. Matthias Geyer (University of Bonn) for providing the Nef protein model structure, Dr. Tzachi Hagai and Stern lab members for helpful comments and discussion.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing Interests

The authors declare that they have no competing interests involved with this study.

Bibliography

1. Hahn, B.H., G.M. Shaw, K.M. De Cock, and P.M. Sharp, *AIDS - AIDS as a zoonosis: Scientific and public health implications.* Science, 2000. **287**(5453): p. 607-614.
2. Van Heuverswyn, F., Y. Li, C. Neel, E. Bailes, B.F. Keele, W. Liu, S. Loul, C. Butel, F. Liegeois, Y. Bienvenue, E.M. Ngolle, P.M. Sharp, G.M. Shaw, E. Delaporte, B.H. Hahn, and M. Peeters, *Human immunodeficiency viruses: SIV infection in wild gorillas.* Nature, 2006. **444**(7116): p. 164.
3. D'Arc, M., et al., *Origin of the HIV-1 group O epidemic in western lowland gorillas.* Proc Natl Acad Sci U S A, 2015. **112**(11): p. E1343-52.
4. Takehisa, J., M.H. Kraus, A. Ayouba, E. Bailes, F. Van Heuverswyn, J.M. Decker, Y. Li, R.S. Rudicell, G.H. Learn, C. Neel, E.M. Ngole, G.M. Shaw, M. Peeters, P.M. Sharp, and B.H. Hahn, *Origin and biology of simian immunodeficiency virus in wild-living western gorillas.* J Virol, 2009. **83**(4): p. 1635-48.
5. Bell, S.M. and T. Bedford, *Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission.* PLoS Pathog, 2017. **13**(7): p. e1006466.
6. Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B.H. Hahn, S. Wolinsky, and T. Bhattacharya, *Timing the ancestor of the HIV-1 pandemic strains.* Science, 2000. **288**(5472): p. 1789-96.
7. Lemey, P., O.G. Pybus, A. Rambaut, A.J. Drummond, D.L. Robertson, P. Roques, M. Worobey, and A.M. Vandamme, *The molecular population genetics of HIV-1 group O.* Genetics, 2004. **167**(3): p. 1059-68.
8. Wertheim, J.O. and M. Worobey, *Dating the Age of the SIV Lineages That Gave Rise to HIV-1 and HIV-2.* Plos Computational Biology, 2009. **5**(5): p. e1000377.
9. Bailes, E., F. Gao, F. Bibollet-Ruche, V. Courgnaud, M. Peeters, P.A. Marx, B.H. Hahn, and P.M. Sharp, *Hybrid origin of SIV in chimpanzees.* Science, 2003. **300**(5626): p. 1713.
10. Leitner, T., M.-C. Dazza, M. Ekwalanga, C. Apetrei, and S. Saragosti, *Sequence Diversity among Chimpanzee Simian Immunodeficiency Viruses (SIVcpz) Suggests that SIVcpz Pts Was Derived from SIVcpz Ptt through Additional Recombination Events.* AIDS research and human retroviruses, 2007. **23**(9): p. 1114-1118.
11. Li, Y., et al., *Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir.* J Virol, 2012. **86**(19): p. 10776-91.
12. Etienne, L., B.H. Hahn, P.M. Sharp, F.A. Matsen, and M. Emerman, *Gene loss and adaptation to hominids underlie the ancient origin of HIV-1.* Cell Host Microbe, 2013. **14**(1): p. 85-92.
13. Longdon, B., M.A. Brockhurst, C.A. Russell, J.J. Welch, and F.M. Jiggins, *The evolution and genetics of virus host shifts.* PLoS Pathog, 2014. **10**(11): p. e1004395.
14. Kuiken, T., E.C. Holmes, J. McCauley, G.F. Rimmelzwaan, C.S. Williams, and B.T. Grenfell, *Host species barriers to influenza virus infections.* Science, 2006. **312**(5772): p. 394-7.
15. Taubenberger, J.K. and J.C. Kash, *Influenza virus evolution, host adaptation, and pandemic formation.* Cell Host Microbe, 2010. **7**(6): p. 440-51.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Parrish, C.R., E.C. Holmes, D.M. Morens, E.-C. Park, D.S. Burke, C.H. Calisher, C.A. Laughlin, L.J. Saif, and P. Daszak, *Cross-species virus transmission and the emergence of new epidemic diseases.* Microbiology and Molecular Biology Reviews, 2008. **72**(3): p. 457-470.
17. Knudsen, B. and M.M. Miyamoto, *A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14512-7.
18. Penn, O., A. Stern, N.D. Rubinstein, J. Dutheil, E. Bacharach, N. Galtier, and T. Pupko, *Evolutionary modeling of rate shifts reveals specificity determinants in HIV-1 subtypes.* PLoS Comput Biol, 2008. **4**(11): p. e1000214.
19. Kuiken, C., B. Foley, T. Leitner, C. Apetrei, B. Hahn, I. Mizrachi, J. Mullins, A. Rambaut, S. Wolinsky, and B. Korber, *HIV sequence compendium 2010*. 2010, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).
20. Sironi, M., R. Cagliani, D. Forni, and M. Clerici, *Evolutionary insights into host–pathogen interactions from mammalian sequence data.* Nature Reviews Genetics, 2015. **16**: p. 224.
21. McLaren, P.J. and M. Carrington, *The impact of host genetic variation on infection with HIV-1.* Nature Immunology, 2015. **16**: p. 577.
22. Kirchhoff, F., *Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses.* Cell Host Microbe, 2010. **8**(1): p. 55-67.
23. Kluge, S.F., D. Sauter, and F. Kirchhoff, *SnapShot: antiviral restriction factors.* Cell, 2015. **163**(3): p. 774-774. e1.
24. Kaslow, R.A., M. Carrington, R. Apple, L. Park, A. Munoz, A. Saah, J.J. Goedert, C. Winkler, S.J. O'Brien, and C. Rinaldo, *Influence of combinations of human major histocompatibility complex genes on the course of HIV–1 infection.* Nature medicine, 1996. **2**(4): p. 405-411.
25. Simon, V., N. Bloch, and N.R. Landau, *Intrinsic host restrictions to HIV-1 and mechanisms of viral escape.* Nature Immunology, 2015. **16**: p. 546.
26. Neil, S.J.D., T. Zang, and P.D. Bieniasz, *Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.* Nature, 2008. **451**(7177): p. 425-U1.
27. Van Damme, N., D. Goff, C. Katsura, R.L. Jorgenson, R. Mitchell, M.C. Johnson, E.B. Stephens, and J. Guatelli, *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**(4): p. 245-252.
28. Jia, B., R. Serra-Moreno, W. Neidermyer, A. Rahmberg, J. Mackey, I.B. Fofana, W.E. Johnson, S. Westmoreland, and D.T. Evans, *Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2.* PLoS Pathog, 2009. **5**(5): p. e1000429.
29. Zhang, F., S.J. Wilson, W.C. Landford, B. Virgen, D. Gregory, M.C. Johnson, J. Munch, F. Kirchhoff, P.D. Bieniasz, and T. Hatziioannou, *Nef proteins from simian immunodeficiency viruses are tetherin antagonists.* Cell Host Microbe, 2009. **6**(1): p. 54-67.
30. McNatt, M.W., T. Zang, T. Hatziioannou, M. Bartlett, I.B. Fofana, W.E. Johnson, S.J. Neil, and P.D. Bieniasz, *Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants.* PLoS Pathog, 2009. **5**(2): p. e1000300.
31. Evans, D.T., R. Serra-Moreno, R.K. Singh, and J.C. Guatelli, *BST-2/tetherin: a new component of the innate immune response to enveloped viruses.* Trends Microbiol, 2010. **18**(9): p. 388-96.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

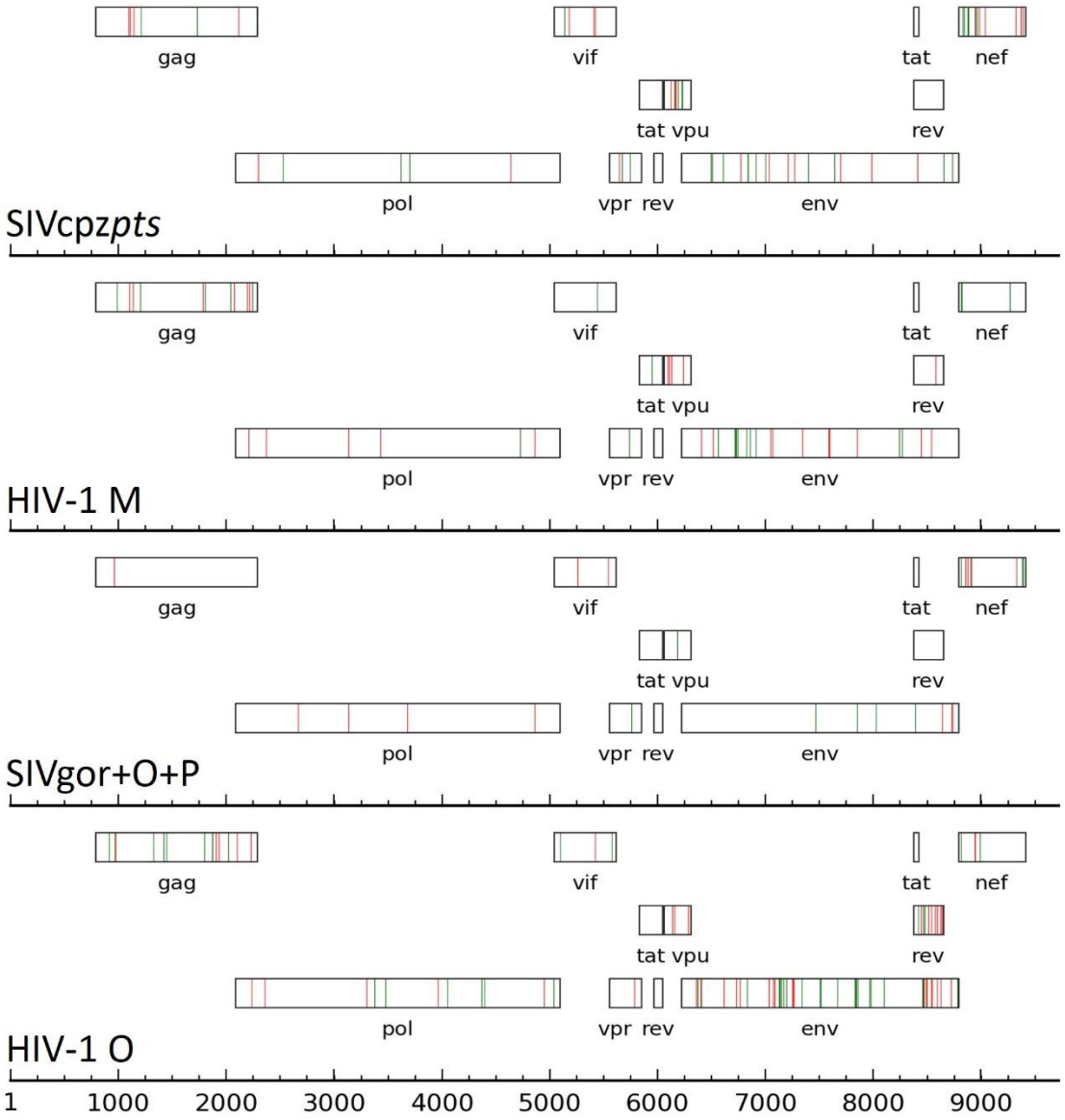
32. Sauter, D., et al., *Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains.* Cell Host Microbe, 2009. **6**(5): p. 409-21.
33. Kluge, S.F., K. Mack, S.S. Iyer, F.M. Pujol, A. Heigele, G.H. Learn, S.M. Usmani, D. Sauter, S. Joas, D. Hotter, F. Bibollet-Ruche, L.J. Plenderleith, M. Peeters, M. Geyer, P.M. Sharp, O.T. Fackler, B.H. Hahn, and F. Kirchhoff, *Nef proteins of epidemic HIV-1 group O strains antagonize human tetherin.* Cell Host Microbe, 2014. **16**(5): p. 639-50.
34. Bego, M.G., L. Cong, K. Mack, F. Kirchhoff, and E.A. Cohen, *Differential Control of BST2 Restriction and Plasmacytoid Dendritic Cell Antiviral Response by Antagonists Encoded by HIV-1 Group M and O Strains.* J Virol, 2016. **90**(22): p. 10236-10246.
35. Mack, K., K. Starz, D. Sauter, S. Langer, F. Bibollet-Ruche, G.H. Learn, C.M. Sturzel, M. Leoz, J.C. Plantier, M. Geyer, B.H. Hahn, and F. Kirchhoff, *Efficient Vpu-Mediated Tetherin Antagonism by an HIV-1 Group O Strain.* J Virol, 2017. **91**(6): p. e02177-16.
36. Sauter, D., S. Hué, S.J. Petit, J.-C. Plantier, G.J. Towers, F. Kirchhoff, and R.K. Gupta, *HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef.* Retrovirology, 2011. **8**(1): p. 103.
37. Lim, E.S., H.S. Malik, and M. Emerman, *Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses.* J Virol, 2010. **84**(14): p. 7124-34.
38. Yang, S.J., L.A. Lopez, C.M. Exline, K.G. Haworth, and P.M. Cannon, *Lack of adaptation to human tetherin in HIV-1 group O and P.* Retrovirology, 2011. **8**(1): p. 78.
39. Gotz, N., D. Sauter, S.M. Usmani, J.V. Fritz, C. Goffinet, A. Heigele, M. Geyer, F. Bibollet-Ruche, G.H. Learn, O.T. Fackler, B.H. Hahn, and F. Kirchhoff, *Reacquisition of Nef-Mediated Tetherin Antagonism in a Single In Vivo Passage of HIV-1 through Its Original Chimpanzee Host.* Cell Host & Microbe, 2012. **12**(3): p. 373-380.
40. Mack, K., *Counteraction of the antiviral factor tetherin by HIV-1 group O*. 2017, Universität Ulm.
41. Geyer, M. and B.M. Peterlin, *Domain assembly, surface accessibility and sequence conservation in full length HIV-1 Nef.* FEBS Lett, 2001. **496**(2-3): p. 91-5.
42. Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono, *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.* Nature, 2003. **424**(6944): p. 99-103.
43. Kato, Y., R. Habas, Y. Katsuyama, A.M. Naar, and X. He, *A component of the ARC/Mediator complex required for TGF beta/Nodal signalling.* Nature, 2002. **418**(6898): p. 641-6.
44. Russell, R.A., J. Smith, R. Barr, D. Bhattacharyya, and V.K. Pathak, *Distinct Domains within APOBEC3G and APOBEC3F Interact with Separate Regions of Human Immunodeficiency Virus Type 1 Vif.* Journal of Virology, 2009. **83**(4): p. 1992-2003.
45. Compton, A.A. and M. Emerman, *Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses.* PLoS Pathog, 2013. **9**(1): p. e1003135.
46. Letko, M., G. Silvestri, B.H. Hahn, F. Bibollet-Ruche, O. Gokcumen, V. Simon, and M. Ooms, *Vif Proteins from Diverse Primate Lentiviral Lineages Use the Same Binding Site in APOBEC3G.* Journal of Virology, 2013. **87**(21): p. 11861-11871.
47. Pery, E., K.S. Rajendran, A.J. Brazier, and D. Gabuzda, *Regulation of APOBEC3 proteins by a novel YXXL motif in human immunodeficiency virus type 1 Vif and simian immunodeficiency virus SIVagm Vif.* J Virol, 2009. **83**(5): p. 2374-81.
48. Feng, Y., T.T. Baig, R.P. Love, and L. Chelico, *Suppression of APOBEC3-mediated restriction of HIV-1 by Vif.* Front Microbiol, 2014. **5**: p. 450.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

49. Letko, M., T. Booiman, N. Kootstra, V. Simon, and M. Ooms, *Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface.* Cell Rep, 2015. **13**(9): p. 1789-99.
50. Hvilsom, C., F. Carlsen, H.R. Siegismund, S. Corbet, E. Nerrienet, and A. Fomsgaard, *Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene.* Genomics, 2008. **92**(5): p. 322-8.
51. Brown, K.E., *Parallel adaptation of CD4 in SIV-endemic Gorilla and Pan lineages*. 2014, University at Albani.
52. Zhou, T., I. Georgiev, X. Wu, Z.Y. Yang, K. Dai, A. Finzi, Y.D. Kwon, J.F. Scheid, W. Shi, L. Xu, Y. Yang, J. Zhu, M.C. Nussenzweig, J. Sodroski, L. Shapiro, G.J. Nabel, J.R. Mascola, and P.D. Kwong, *Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01.* Science, 2010. **329**(5993): p. 811-7.
53. Yamaguchi-Kabata, Y. and T. Gojobori, *Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes.* J Virol, 2000. **74**(9): p. 4335-50.
54. Wain, L.V., E. Bailes, F. Bibollet-Ruche, J.M. Decker, B.F. Keele, F. Van Heuverswyn, Y. Li, J. Takehisa, E.M. Ngole, G.M. Shaw, M. Peeters, B.H. Hahn, and P.M. Sharp, *Adaptation of HIV-1 to its human host.* Mol Biol Evol, 2007. **24**(8): p. 1853-60.
55. Peeters, M., A. Gueye, S. Mboup, F. Bibollet-Ruche, E. Ekaza, C. Mulanga, R. Ouedrago, R. Gandji, P. Mpele, G. Dibanga, B. Koumare, M. Saidou, E. Esu-Williams, J.P. Lombart, W. Badombena, N. Luo, M. Vanden Haesevelde, and E. Delaporte, *Geographical distribution of HIV-1 group O viruses in Africa.* AIDS, 1997. **11**(4): p. 493-8.
56. Leoz, M., F. Feyertag, A. Kfutwah, P. Mauclere, G. Lachenal, F. Damond, F. De Oliveira, V. Lemee, F. Simon, D.L. Robertson, and J.C. Plantier, *The Two-Phase Emergence of Non Pandemic HIV-1 Group O in Cameroon.* PLoS Pathog, 2015. **11**(8): p. e1005029.
57. Hasegawa, M., H. Kishino, and T. Yano, *Dating of the human-ape splitting by a molecular clock of mitochondrial DNA.* J Mol Evol, 1985. **22**(2): p. 160-74.
58. Ruvolo, M., *Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets.* Mol Biol Evol, 1997. **14**(3): p. 248-65.
59. Loytynoja, A. and N. Goldman, *An algorithm for progressive multiple alignment of sequences with insertions.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(30): p. 10557-10562.
60. Guindon, S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Syst Biol, 2010. **59**(3): p. 307-21.
61. Schultz, A.K., M. Zhang, I. Bulla, T. Leitner, B. Korber, B. Morgenstern, and M. Stanke, *jpHMM: improving the reliability of recombination prediction in HIV-1.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W647-51.

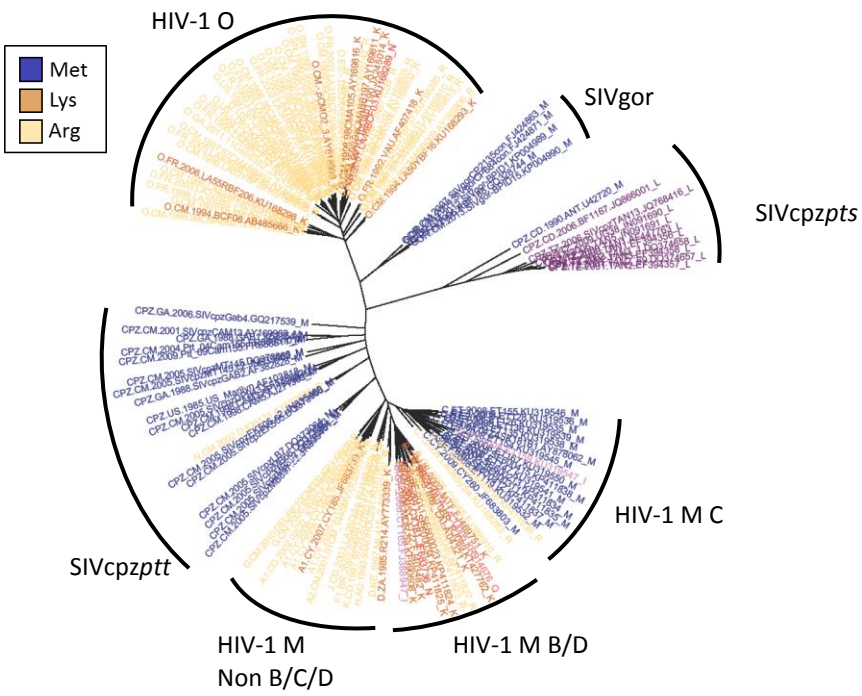
bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figures and Tables



Supplementary Figure 1. Proposed rate deceleration patterns in prominent branches shown along the genome. Data shown in HXB2 coordinates.

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



494

495

496 **Supplementary Figure 2. Residues distribution for Gag_{30} along the HIV-1/SIV phylogeny.** This

497 position exhibits a pattern more similar to "content shift" than to "rate shift".

498

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 1. Rate shifts as *percentage from total rate shifts for prominent branches*, for rate decelerating sites (upper) and rate accelerating sites (lower), colored by intensity.

DEC Branch/protein	Env	Gag	Pol	Rev	Tat	Nef	Vif	Vpr	Vpu
Group M	39%	21%	21%	4%	0%	0%	0%	0%	14%
Group O	43%	11%	11%	20%	0%	4%	2%	2%	7%
SIVgor+P+O	21%	7%	29%	0%	0%	29%	14%	0%	0%
SIVcpz*pts*	26%	15%	7%	0%	0%	22%	11%	4%	15%
ACC Branch/protein	**Env**	**Gag**	**Pol**	**Rev**	**Tat**	**Nef**	**Vif**	**Vpr**	**Vpu**
Group M	43%	22%	4%	0%	4%	13%	4%	4%	4%
Group O	55%	18%	12%	6%	0%	4%	4%	0%	2%
SIVgor+P+O	40%	0%	0%	0%	0%	40%	0%	10%	10%
SIVcpz*pts*	35%	10%	10%	0%	0%	23%	3%	6%	13%

Supplementary Table 2. Maximum log-likelihood (LL) values for the analysis of the nine HIV-1/SIVcpz/SIVgor proteins under the rate shift and null models.

Protein	Rate Shift Model LL	Null Model LL	2ΔLL	P-value (χ^2_3)
Gag	-29,478.3	-29,671.4	386	$<10^{-50}$
Pol	-40,813.1	-41,097.2	568	$<10^{-100}$
Vif	-12,677.6	-12,756.3	157	$<10^{-30}$
Vpr	-5,328.1	-5,355.76	55	$<10^{-10}$
Tat	-8,918.23	-8,947.61	59	$<10^{-10}$
Rev	-12,662	-12,733.6	143	$<10^{-30}$
Vpu	-10,548.7	-10,664.1	231	$<10^{-50}$
Env	-92,347.6	-93,028.8	1362	$<10^{-100}$
Nef	-15,611.4	-15,804.1	385	$<10^{-50}$

bioRxiv preprint doi: https://doi.org/10.1101/190769; this version posted January 18, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary and Source Files

Supplementary File 1.

HIV/SIV sites identified as rate shifting for dataset (i). Positions are provided in HXB2 reference sequence coordinates. Branch number field refer to the branch number outputted by RASER [18]; for ease of reading we provide branch labels for most branches.

Supplementary File 2.

HIV/SIV sites identified as rate shifting for dataset (ii). Positions are provided in HXB2 reference sequence coordinates. Branch number field refer to the branch number outputted by RASER [18]; for ease of reading we provide branch labels for most branches.