Gene Loss and Adaptation to Hominids Underlie the Ancient Origin of HIV-1

Lucie Etienne,[1] Beatrice H. Hahn,[4] Paul M. Sharp,[5] Frederick A. Matsen,[2] and Michael Emerman[1,3,*]
[1]Division of Human Biology
[2]Division of Public Health Sciences
[3]Division of Basic Science
Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[4]Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK
*Correspondence: memerman@fhcrc.org
http://dx.doi.org/10.1016/j.chom.2013.06.002

SUMMARY

HIV-1 resulted from cross-species transmission of SIVcpz, a simian immunodeficiency virus that naturally infects chimpanzees. SIVcpz, in turn, is a recombinant between two SIV lineages from Old World monkeys. Lentiviral interspecies transmissions are partly driven by the evolution and capacity of viral accessory genes, such as *vpx, vpr*, and *vif*, to antagonize host antiviral factors, such as SAMHD1 and the APOBEC3 proteins. We show that *vpx*, which in other lentiviruses antagonizes *SAMHD1*, was deleted during the creation of SIVcpz. This genomic deletion resulted in the reconstruction of the overlapping *vif* gene by "overprinting," creating a unique *vif* that overlaps in its 3′ end with the *vpr* gene and can antagonize hominid *APOBEC3s*. Moreover, passage of SIVs through chimpanzees facilitated the subsequent adaptation of HIV-1 to humans. Thus, HIV-1 originated through a series of gene loss and adaptation events that generated its chimpanzee precursor and lowered the species barrier to human infection.

INTRODUCTION

HIV-1 is the result of cross-species transmissions of simian immunodeficiency viruses (SIVs) from African apes to humans. SIVcpz strains from chimpanzees (*Pan troglodytes*) were transmitted on at least two occasions to humans, including the cross-species transmission of the precursor of HIV-1 group M that spawned the current AIDS pandemic (Keele et al., 2006; Sharp and Hahn, 2011). SIVcpz, in turn, originated from interspecies transmissions and recombination events involving the ancestors of at least two distant SIV lineages: SIVrcm from red-capped mangabeys (RCMs) and SIVmus/mon/gsn from *Cercopithecus* monkeys (Bailes et al., 2003). Although the adaptation of lentiviruses from chimpanzees to humans has been described (Kirchhoff, 2010), the adaptive processes involved in the transmission of SIVs from monkeys to chimpanzees, which underlie the ultimate origin of HIV-1, are not well understood.

Host susceptibility to viral infections and the likelihood of lentiviral transmission from one primate species to another are partially governed by the antiviral proteins produced by the innate immune system of the host. These proteins, also called restriction factors, inhibit different stages of lentiviral replication and are usually counteracted in a species-specific manner by viral accessory proteins (Duggal and Emerman, 2012). One well-described antagonism is the degradation of the host cytidine deaminase APOBEC3G protein (A3G) by the viral infectivity protein Vif (Bishop et al., 2008; Mangeat et al., 2003; Sheehy et al., 2002). In the absence of Vif, A3G is packaged into assembling virions and transferred to target cells, where it produces hypermutation in the viral genome. However, during infection, Vif interacts with A3G and recruits it to a ubiquitin ligase complex to target A3G for proteasomal degradation, which prevents the encapsidation of A3G and eventually allows viral replication in the new target cells. A more recently identified virus-host antagonism is the degradation of the host protein SAMHD1 by the accessory proteins Vpx and Vpr to allow the virus to efficiently infect myeloid and resting T cells (Baldauf et al., 2012; Hrecka et al., 2011; Laguette et al., 2011; Lim et al., 2012). The capacity to antagonize the host SAMHD1 was acquired by the *vpr* gene during primate lentiviral evolution. Subsequently, the recombination/duplication of the *vpr* gene led to the acquisition of a *vpx* gene in SIVrcm and SIVsmm from sooty mangabeys (Lim et al., 2012). The two viral lineages that recombined to give rise to SIVcpz have the capacity to degrade their respective SAMHD1 proteins: SIVrcm uses its Vpx protein, whereas members of the SIVmus/mon/gsn lineage use their Vpr protein (Lim et al., 2012). In contrast, neither HIV-1 nor SIVcpz, its immediate precursor, have the ability to degrade SAMHD1 because they do not encode a *vpx* gene, and their Vpr protein does not antagonize SAMHD1 (Laguette et al., 2011; Lim et al., 2012). Thus, one seemingly important function to antagonize a host restriction factor was lost during the process of primate lentivirus adaptation from monkeys to hominids, but the mechanism and the reason for this loss remain unknown.

Here, we analyzed the viral genomic reorganization and functional consequences that occurred during the transmission of lentiviruses from Old World monkeys (OWMs) to hominids in order to understand the selective pressures that led to the ultimate origin of HIV-1 strains. We found that the *vpx* gene was entirely lost during the birth of SIVcpz and led to the absence of *vpx* in HIV-1. Furthermore, this loss was associated with the reconstruction of the overlapping *vif* gene. This Vif protein, which is unique to SIVcpz and its descendants, gained the function to fully antagonize the hominid A3G proteins. Finally, we found that chimpanzees represented a means for a primate lentivirus to adapt to hominids, which was likely essential for their subsequent transmission to humans. Our study elucidates how the HIV-1 lineage that led to the emergence of a pandemic virus in the human population had its ultimate origin in a monkey-to-hominid cross-species transmission that involved the loss of a viral gene, the creation of a distinctive 3′ end region to an existing gene, and subsequent adaptation in chimpanzees.

RESULTS

The Absence of *vpx* in HIV-1 Results from the Loss of the Entire Gene during the Genesis of SIVcpz

Previous phylogenetic analyses of the SIVcpz genome showed that *vif* and *vpr*, among other genes, originated from an ancient SIVrcm-like virus, whereas *vpu* and *env* genes came from an ancient SIVmus/mon/gsn-like strain (Sharp and Hahn, 2011), and we confirmed these conclusions by analyses that included all of the SIV lineages characterized to date (data not shown). However, SIVrcm encodes for a *vpx* gene between *vif* and *vpr*, whereas *vpx* is absent in SIVcpz (Figure 1A). Thus, we asked how and why the *vpx* gene came to be absent in SIVcpz. We first considered the possibility that the ancestor of the modern-day SIVrcm lacked the *vpx* gene when it crossed to chimpanzees. Therefore, we analyzed an alignment of the surrounding regions of *vpx* to look for any evidence of a recent *vpx* gene transfer to SIVrcm (details of the hypotheses and analyses are in Figure S1 available online). The absence of recombination marks and the phylogenetic tree topologies revealed a lack of evidence for any *vpx* gene transfer between SIVsmm and SIVrcm, showing that *vpx* was acquired by the SIVrcm and SIVsmm lineages before they diverged (Figure S1). This indicates that the SIVrcm-like strain, which recombined with a SIVmus-like strain to give rise to SIVcpz, encoded a *vpx* gene between *vif* and *vpr*. Thus, SIVcpz lost the *vpx* gene from the ancestral SIVrcm-like virus.

We then considered two main scenarios for how the *vpx* gene was lost during the origin of SIVcpz (Figure 1A). In scenario 1, recombination occurred between the paralogous genes *vpr* and *vpx*, leading to just one gene in SIVcpz, where the 5′ end derives from SIVrcm *vpx* and the 3′ end derives from SIVrcm *vpr*. In scenario 2, the entire *vpx* was deleted in SIVcpz, but the *vpr* gene from SIVrcm remained completely intact (Figure 1A). The difference between these two scenarios is important because Vpr and Vpx perform different functions (Ayinde et al., 2010), and recombination between the genes (scenario 1) would imply that some of the functional domains of both genes were conserved, whereas direct deletion of *vpx* (scenario

2) would imply that only the functions of one gene, *vpr*, were conserved.

We found multiple lines of evidence indicating that scenario 1 is not plausible, but scenario 2 is very likely. First, phylogenetic analysis of Vpr and Vpx from diverse SIVs showed that SIVcpz Vpr is closely related to SIVrcm Vpr, and both cluster distantly from SIVrcm Vpx (Figure 1B). Second, if SIVcpz Vpr was a recombinant between SIVrcm Vpx and Vpr (scenario 1), one would expect the N-terminal (N-ter) region of SIVcpz Vpr to show more similarity to Vpx than to Vpr (Figure 1A). However, when we aligned sequences of SIVcpz Vpr with sequences of SIVrcm Vpr and Vpx, we found that this was not true, and instead the N-ter of SIVrcm and SIVcpz Vpr shared higher similarity (the SIVcpz *vpr* gene has 54% identity with SIVrcm *vpr* versus 39% with SIVrcm *vpx*; Figure 1C). Third, we found neither evidence of ancient recombination events within SIVcpz *vpr*, nor any remnant genomic region of SIVrcm *vpx* in SIVcpz (see Experimental Procedures). These results exclude the first scenario of recombination between *vpr* and *vpx*, and favor a scenario in which the entire *vpx* gene was lost in SIVcpz.

The Loss of *vpx* Led to the Creation of a *vif* with a Unique 3′ Terminal Region by "Overprinting"

To determine the extent of the deletion that led to the loss of *vpx*, the impact on the overlapping genes, and the associated genomic modifications, we analyzed the region spanning *vif*, *vpx*, and *vpr*. In SIVrcm, the 5′ end of *vpx* overlaps with the 3′ end of *vif* by ~160 bp (Figure 2). However, in SIVcpz, we found that the loss of *vpx* led to the loss of this entire overlapping region, including the stop codon for *vif* (Figure 2). Thus, we speculated that the 3′ end truncation of the *vif* gene led to its reconstruction in an alternate reading frame of the *vpr* gene by a mechanism called "overprinting." Overprinting is the process by which a sequence that originally encodes for only one protein undergoes modifications leading to an additional second open reading frame (ORF) (Keese and Gibbs, 1992). Indeed, we found that during the recombination events that generated SIVcpz, the 3′ end of *vif* was reconstituted by overprinting of the 5′ end of SIVrcm *vpr* (ORF 2; Figure 2). Here, in frame 3 of SIVrcm (shown in dark purple in Figure 2, top), a pre-existing stop codon within the SIVrcm *vpr* gene (but not in-frame with SIVrcm *vpr*) served as the stop codon for SIVcpz *vif* after the deletion of *vpx* from SIVrcm (Figures 2 and S2). Hence, in SIVcpz, the 3′ end of *vif* (the overprinting region) overlaps with the 5′ end of *vpr* by ~61 bp in an alternative reading frame (dark purple, Figure 2). Furthermore, upstream of the overprinting region of SIVcpz *vif*, a sequence of 60–75 bp in SIVcpz was created (light purple, Figure 2) for which we could find no homology with any genomic region of SIVrcm or to any other nucleotide sequence present in any sequenced primate genome or lentivirus (see Experimental Procedures). These Vif sequences are unique to SIVcpz and its related strains, SIVgor and HIV-1. Importantly, this fragment harbors the "cullin box," which includes the PPLP motif and surrounding residues that are highly conserved in HIV-1 and are necessary for HIV-1 Vif to efficiently degrade human A3G (Bergeron et al., 2010; Donahue et al., 2008; Walker et al., 2010). Thus, SIVcpz Vif and therefore HIV-1 Vif acquired a unique C-terminal (C-ter) domain with sequences important for their protein function during the loss of Vpx.
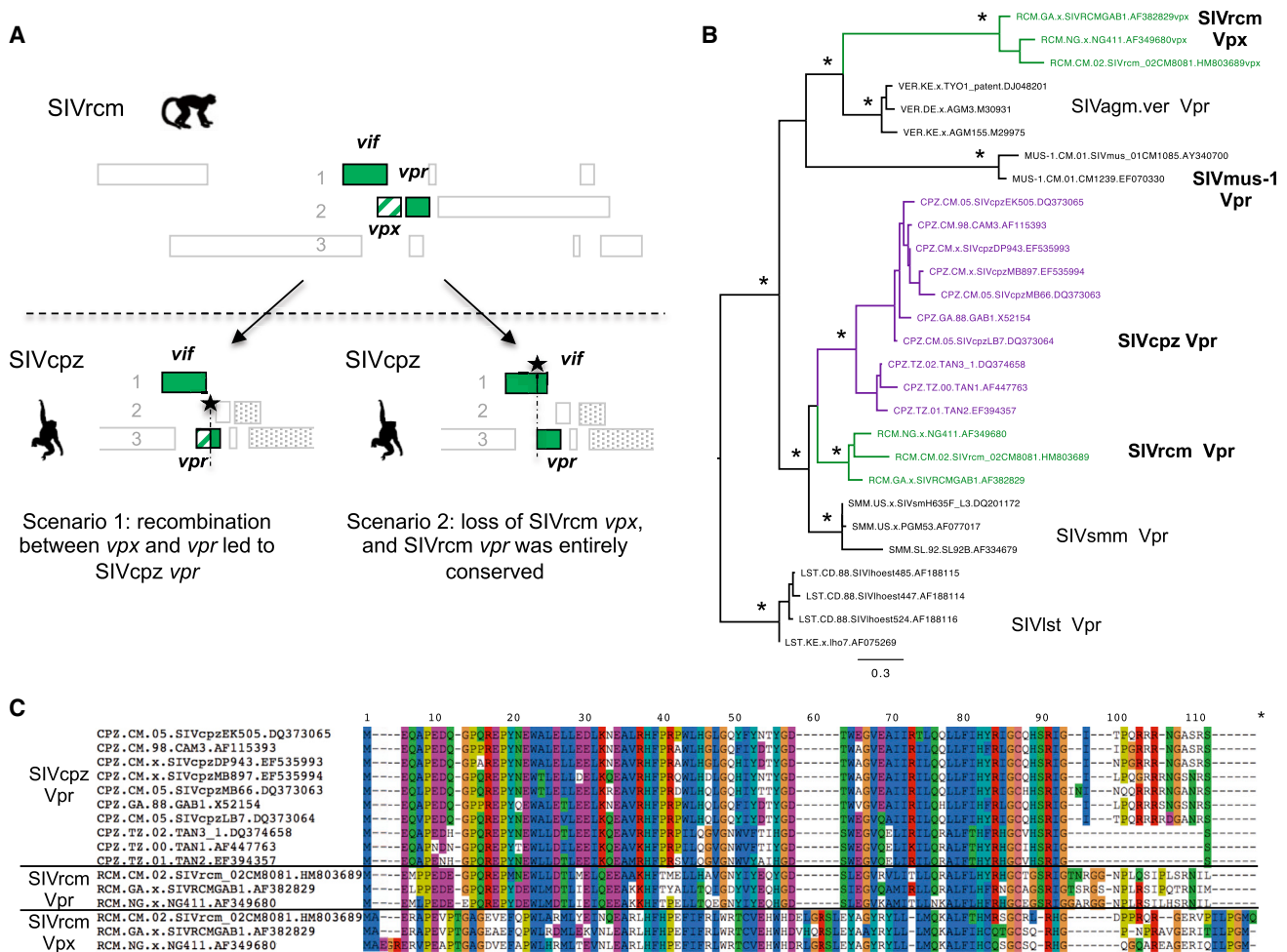
Figure 1. The Entire *vpx* Gene Was Lost in the Lineage that Gave Rise to SIVcpz, SIVgor, and HIV-1

(A) Two possible scenarios leading to the absence of *vpx* in SIVcpz. Only the region of interest spanning *vif*, *vpx*, and *vpr* is detailed. The horizontal dashed line represents the transmission from RCMs (top) to chimpanzees (bottom). The potential recombination breakpoints of interest are represented by plain stars and the *vpx* gene is highlighted by striped lines. In the SIVcpz genome, the dotted regions correspond to the *vpu* gene and the *env* gene that originated from SIVmus/mon/gsn-like viruses. The numbers correspond to the three reading frames.

(B) Vpr from SIVcpz is more closely related to SIVrcm Vpr than to SIVrcm Vpx or SIVmus Vpr. The phylogenetic analysis was performed from an alignment of full-length Vpr and Vpx from various SIVs (trimmed alignment of 76 amino acids). The name and accession number of each SIV are at the tip of each branch. Sequences were retrieved from http://www.hiv.lanl.gov. The asterisks show bootstrap values superior to 85%.

(C) The N-ter region of SIVcpz Vpr is closely related to the N-ter region of SIVrcm Vpr. The amino acid alignment of full-length SIVcpz Vpr, SIVrcm Vpr, and SIVrcm Vpx is shown. See also Figure S1.

SIVrcm *vif* Adapted to Efficiently Antagonize Chimpanzee A3G and A3D, but SAMHD1 Antagonism Was Not Retained

The capacity of Vif to antagonize the host A3G is a feature that is conserved throughout SIVs and primates; however, this antagonism is often species specific (Compton and Emerman, 2013). We hypothesized that the loss of *vpx* during the birth of SIVcpz was driven by selection for changes in the overlapping reading frame that encodes Vif, i.e., the ability of Vif to antagonize chimpanzee A3G was of greater importance than the need to retain Vpx for the purpose of SAMHD1 antagonism. In support of this hypothesis, we found that neither SIVrcm Vpx nor SIVrcm Vpr was able to degrade the chimpanzee SAMHD1 (Figure 3A), although, as expected (Lim et al., 2012), SIVrcm Vpx, but not SIVrcm Vpr, had the capacity to degrade SAMHD1 from RCMs

(Figure 3A). Thus, the absence of SAMHD1 antagonism presumably had little consequence for the initial transmission of SIV from monkeys to chimpanzees.

We therefore tested whether the deletion in *vpx* was important for the adaptation of *vif* to antagonize chimpanzee A3G. First, we tested the ability of SIVrcm Vif to antagonize chimpanzee A3G. We found that although Vif from SIVrcm was potent at antagonizing its own host A3G (Figure 3B, top in plain green), it was only able to partially counteract the antiviral effects of chimpanzee A3G (16% rescue versus 100% for SIVcpz Vif; Figure 3B, bottom plain green versus purple). Degradation assays further confirmed that SIVrcm and SIVcpz Vifs efficiently degraded RCM and chimpanzee A3Gs, respectively, whereas neither Vif degraded the heterologous A3G (Figures 3C and 3D). Thus, SIVrcm Vif needed to adapt
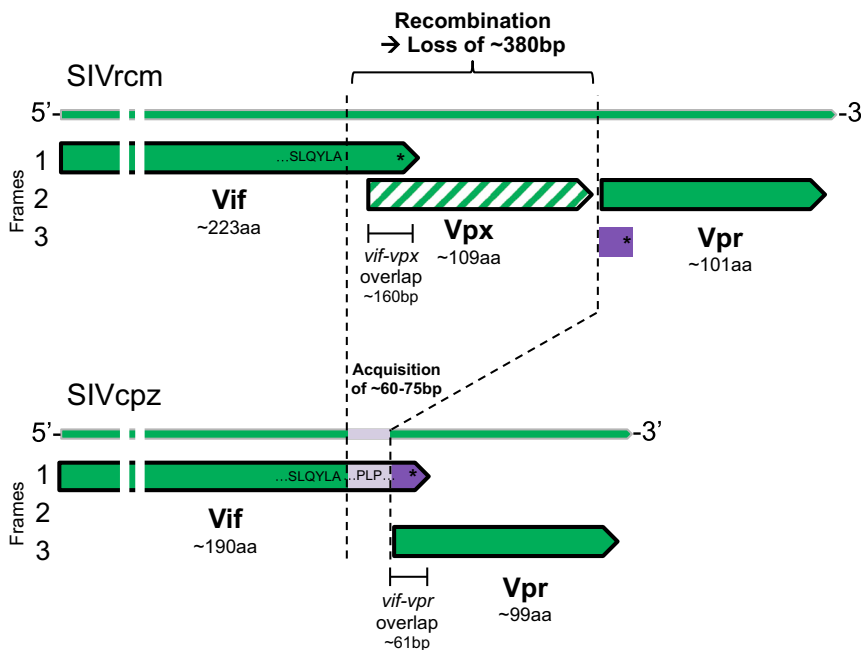
Figure 2. The Loss of *vpx* Had Major Consequences for *vif*

The 3′ end of SIVrcm *vif* was reconstructed when SIVrcm *vpx* was lost during the birth of SIVcpz. Representations of SIVrcm (top) and SIVcpz (bottom) genomes in the region spanning the *vif*, *vpx* (striped arrow), and *vpr* ORFs. The approximate length of each protein is given. The reading frames (1–3) are given on the left and the proteins are represented by large plain arrows. Green stands for proteins and nucleotides related to SIVrcm, dark purple stands for sequences that were not expressed in SIVrcm but were in an ORF in SIVcpz (overprinting region), and light purple stands for new amino acids in SIVcpz. Asterisks are stop codons. Gene overlaps and important nucleotide and amino acid motifs or regions are shown. The dashed lines represent breakpoints. The white lines in the 5′ region of the arrows indicate regions that were cut for representation purposes only. aa, amino acids. See also Figure S2.

to efficiently antagonize chimpanzee A3G and gain full infectivity.

We found that, alone, the unique C-ter part of SIVcpz Vif that was reconstructed upon the loss of *vpx* was not sufficient to rescue the infection in the presence of chimpanzee A3G (Figure 3B, chimera1). However, this C-ter domain of SIVcpz Vif was essential for the protein function, as its replacement with the C-ter part of SIVrcm Vif led to a chimera that had no activity against A3G (Figure 3B, chimera2). Changes in SIVrcm Vif to any of the conserved N-ter motifs from SIVcpz/HIV-1 Vif that are essential for Vif function (reviewed in Malim and Bieniasz, 2012; Figure S3A) led to chimeric Vifs that lost their activity against chimpanzee A3G but retained their activity against RCM A3G (Figure S3B). Thus, the domains in SIVrcm Vif that are required for the degradation of chimpanzee A3G and RCM A3G are distinct and nonlinear, and changes at both the N-ter and C-ter of SIVrcm Vif were needed to fully gain chimpanzee A3G antagonism.

We also investigated whether SIVrcm Vif needed to adapt to antagonize other genes from the chimpanzee APOBEC3 family. We found that SIVrcm Vif has the capacity to antagonize both chimpanzee A3F and A3H (Figure 3E). Therefore, chimpanzee A3F and A3H did not represent barriers for SIVrcm-like strains to jump to chimpanzees. On the other hand, SIVrcm Vif did not antagonize chimpanzee A3D (Figure 3E). Chimpanzee A3D is particularly interesting because it gained increased antiviral activity against some lentiviruses due to positive selection in this gene in the chimpanzee-bonobo lineage (Duggal et al., 2011). Thus, SIVrcm Vif needed to adapt to fully antagonize chimpanzee A3G and A3D, and the deletion of *vpx* may have been required for these changes to take place.

Chimpanzees as a "Passage" for Lentiviruses to Human Infection

The only lentiviruses that are known to have directly transferred from OWMs to humans are strains of SIVsmm from sooty mangabeys, which gave rise to HIV-2 (Gao et al., 1992; Santiago et al., 2005). Interestingly, Vif from SIVsmm is pre-equipped to counteract human A3G (Compton and Emerman, 2013). Because SIVrcm and SIVmus never crossed to humans, as opposed to their recombinant progeny, SIVcpz, we hypothesized that passage through chimpanzees was a determinant step for *vif* to adapt to humans. Thus, we sought to determine whether adaptation of SIV Vif to human A3G antagonism occurred during the transfer from chimpanzees to humans or during the adaptation to chimpanzees. We found that SIVcpz Vif rescued the infection to levels similar to those observed for HIV-1 Vif in the presence of human A3G, and that it could efficiently degrade the host protein (Figure 4A, left panel), consistent with previous studies (Gaddis et al., 2004). On the other hand, we found that SIVrcm Vif, as well as SIVmus Vif, had very little capacity to rescue the infection (<10%) in the presence of human A3G (Figure 4A, left panel) as compared with SIVcpz or HIV-1 Vifs (p < 0.0001). Hence, human A3G is a hurdle for SIVrcm and SIVmus, but not for SIVcpz. Furthermore, SIVrcm and SIVmus Vifs were better at antagonizing chimpanzee A3G than human A3G (Figure 4A, comparing right and left panels; 3-fold difference in infectivity, p < 0.005). Thus, the adaptive hurdle from monkeys to chimpanzees appears to be lower than the barrier directly from monkeys to humans for SIVs, such as SIVrcm or SIVmus, to overcome A3G antagonism. Therefore, the chimpanzee-adapted Vif that was pre-equipped to antagonize human A3G may have been a path to transmission to humans.

DISCUSSION

In this work, we studied the deep origins of HIV-1 by investigating the gene loss and adaptations that occurred during and after lentiviruses were transferred from monkeys to chimpanzees to create SIVcpz. Specifically, we found that (1) the SIVrcm *vpx* gene was lost in its entirety upon adaptation to chimpanzees; (2) the loss of *vpx* was associated with the creation of a unique *vif* region by overprinting; (3) Vif adapted to antagonize

chimpanzee APOBEC3 proteins, including A3G and A3D; and (4) the chimpanzee-adapted lentivirus was more efficient than its monkey ancestors at antagonizing human restriction factors such as A3G (Figure 4B). Thus, lentiviral gene loss and adaptations in the chimpanzee host were at the origin of the human HIV-1 pandemic.

Although SAMHD1 antagonism by the *vpx* or *vpr* genes is conserved in many lentiviruses (Lim et al., 2012), we found that this function is not strictly necessary for viral adaptation to hominids. It was previously hypothesized that the ancestral SIVcpz had a *vpx* gene and was transmitted from chimpanzees to RCMs and mandrills, giving rise to SIVrcm(Vpx+) and SIVmnd2(Vpx+), respectively, and that subsequently SIVcpz lost its *vpx* gene in its natural host (Zhang et al., 2012). However, this scenario is not plausible because the ancestor of SIVrcm and SIVsmm acquired the *vpx* gene prior to their divergence and prior to the jump of SIVs from RCMs to chimpanzees (Figure S1).

It is remarkable that a viral antagonist to a host protein could be lost from the viral genome, since one would have expected the adaptation of Vpx to counteract the new host SAMHD1 (Lim et al., 2012). However, the selective pressure during the cross-species jump from OWMs to chimpanzees may have favored the maintenance of a virus in which a poorly active gene was lost by recombination but a critical function in an over-lapping gene was restored or gained. Vif adaptation to chimpanzee APOBEC3s may have been more critical than Vpx adaptation to chimpanzee SAMHD1. This suggests that the A3 proteins are a more potent selective force in the transmission of lentiviruses than SAMHD1. Finally, it is possible that SIVcpz acquired functions to compensate for the absence of Vpx-driven antagonism of SAMHD1, and that these adaptations have increased pathogenicity in the HIV-1 lineage.

The loss of *vpx* together with additional modifications led to a unique viral *vif* gene. However, the origin of the region that includes the cullin box in SIVcpz/HIV-1 Vif is unknown. It is possible that these sequences arose during the *vif* reconstruction associated with the species jump, or they may have also been acquired later. Although SIVcpz, SIVgor, and HIV-1 Vifs harbor a cullin box with a highly conserved PLP motif, most SIV Vifs harbor a very divergent cullin box that has a distinct evolutionary history, which explains why HIV-1 has a distinctive way to antagonize human A3G compared with HIV-2 and other SIVs (Barraud et al., 2008; Gaur and Strebel, 2012). Although SIVsmm, which crossed to humans, harbors a Vif with such a divergent C-ter domain, its Vif and Vpx were pre-equipped to antagonize human A3G and SAMHD1, respectively (Compton and Emerman, 2013; Hrecka et al., 2011). Whether the unique cullin box in Vif from the SIVcpz/HIV-1 lineage has a distinctive role remains to be addressed, but in any case, the passage of the SIVrcm/SIVmus-like recombinant to hominids was not as easy as the passage of SIVsmm to humans. Furthermore, we found that mutations in both the N- and C-ter domains of Vif were necessary to gain full chimpanzee A3G antagonism. Since an SIVrcm Vif with only the C-ter reconstruction of SIVcpz Vif (analogous to chimera1 in Figure 3) is poorly active against chimpanzee APOBEC3 proteins, the deletion of *vpx* may have put the intermediate virus into a fitness valley from which it could have recovered only with additional N-ter mutations in Vif. However,

it is also possible that an ancestral version of SIVrcm Vif already harbored a Vif with an N-ter domain that directly provided full activity to Vif after its C-ter reconstruction in SIVcpz.

Due to species specificities in virus-host interactions and antagonisms, lentiviruses need to adapt to the new host proteins before they can efficiently infect a new species. This jump is easier to make when the virus is pre-equipped to antagonize the recipient species' restriction factors (e.g., human A3G is not a barrier for SIVcpz). However, we further confirmed that human A3G cannot be counteracted by SIVs from most OWMs, with the exception of SIVsmm, which crossed to humans on multiple occasions (Compton and Emerman, 2013; Gao et al., 1992; Santiago et al., 2005). Thus, human A3G appears to pose a major hurdle for SIVs from most OWMs, limiting their transfer to humans. On the other hand, viruses such as SIVrcm and SIVmus have some activity against chimpanzee A3G, which may have played a key role in enabling their jump to apes by allowing an initial poorly efficient viral infection followed by viral adaptation in the new host. Since A3G has been under strong positive selection in the human genome after the divergence of *Pan* and *Homo* due to an unknown selective pressure (Sawyer et al., 2004), it is possible that this evolution led to a variation in human A3G that is more poorly recognized by monkey SIVs than chimpanzee A3G. Therefore, the chimpanzee host may have constituted an intermediary in the adaptive processes that allowed certain OWM lentiviruses to infect humans. Although adaptation in the chimpanzee host does not provide adaptation to all human proteins (e.g., SIVcpz is not pre-equipped to antagonize human Tetherin; Sauter et al., 2009), we propose that SIV adaptation to chimpanzee restriction factors reduced both the number and size of the hurdles for cross-species transmission to humans, which favored a successful viral emergence in the human population.

In summary, it is possible to trace back many of the ancient genetic events in the evolution of primate lentiviruses that ultimately led up to the emergence of HIV-1 (Figure 4B), and the adaptations that occurred during the cross-species transmissions leading to SIVcpz in chimpanzees reveal the functional origins of the pandemic HIV-1 in humans.

EXPERIMENTAL PROCEDURES

Plasmids

SAMHD1, Vpr, and Vpx expression plasmids were previously described (Lim et al., 2012). Apobec3 and recombinant HIV-1 proviral plasmids were constructed as detailed in the Supplemental Experimental Procedures.

Transfection and Western Blot Analysis for the SAMHD1-Vpr/Vpx Study

293T cells were transfected with 100 ng of SAMHD1 expression plasmid with or without 100 ng of Vpr/Vpx constructs using TransIT-LT1 transfection reagent (Mirus Bio). The amounts of transfected Vpr/Vpx and SAMHD1 expression plasmid were normalized for similar protein expression, but the total quantity of transfected DNA was maintained constant. Cells were harvested 48 hr posttransfection for western blot analysis as previously described (Lim et al., 2012). Details of the antibodies are provided in Supplemental Experimental Procedures.

Single-Round Viral Infectivity Assay and Western Blot Analysis for the A3-Vif Study

293T cells were transfected with 400 ng of A3 plasmid or an empty pcDNA3.1 vector, 200 ng of L-VSV-G (the fusogenic envelope G glycoprotein of the vesicular stomatitis virus used for pseudotyping), and 600 ng of proviral plasmid
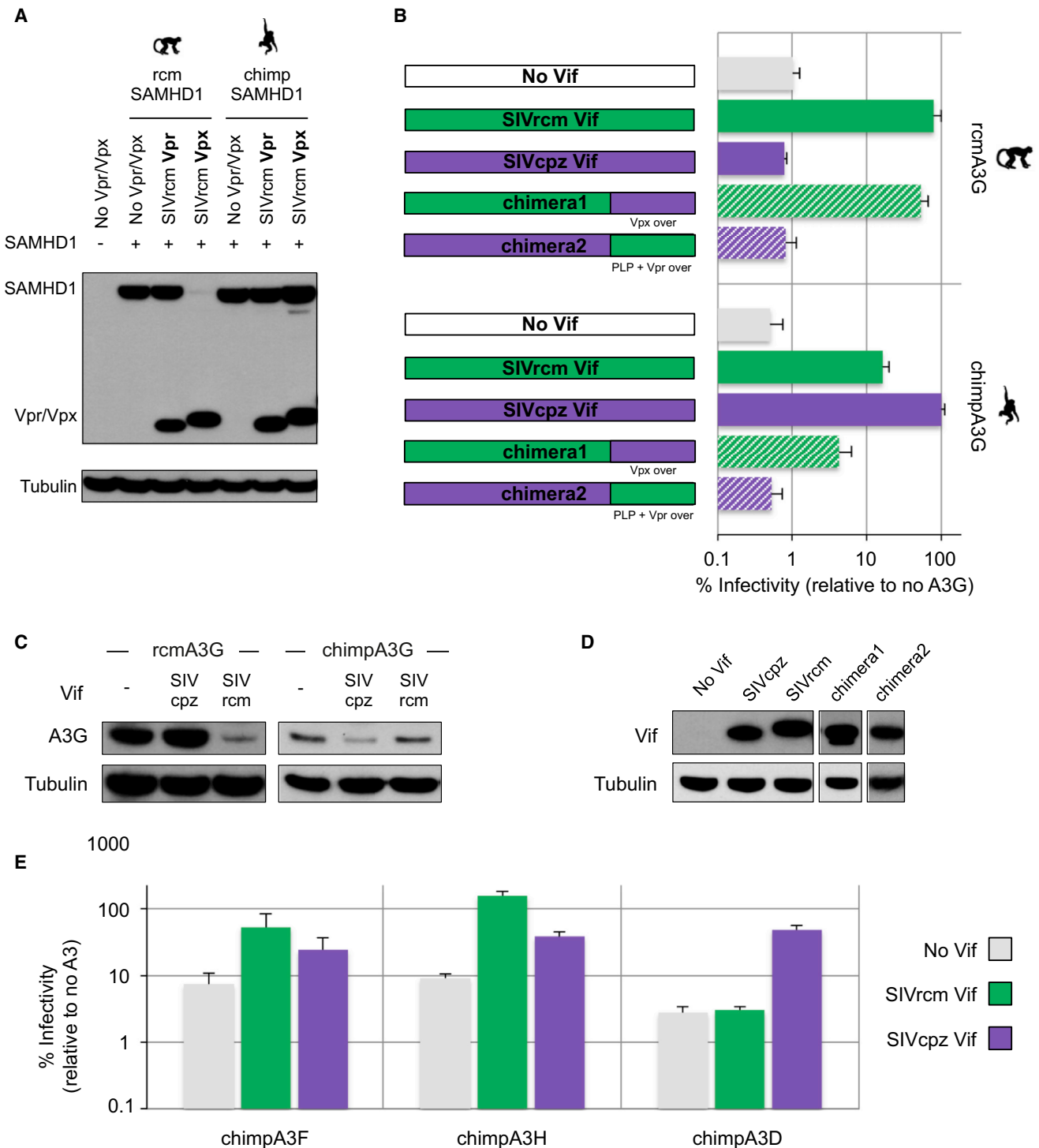
Figure 3. SIVrcm Vif Adaptation to the Hominid APOBEC3 proteins

(A) Vpx and Vpr from SIVrcm do not degrade chimpanzee SAMHD1. The ability of SIVrcm to degrade SAMHD1 was assayed by western blot analyses of hemagglutinin (HA)-tagged SAMHD1 from RCM (rcm) or chimpanzee (chimp) (+, presence; −, empty plasmid) cotransfected with or without 3xFLAG-tag Vpr or Vpx from SIVrcmNG411. Tubulin was probed as a loading control.

(B) SIVrcm Vif has some activity against chimpanzee A3G to rescue viral infection, but adaptions were needed for full antagonism. Single-round infectivity assays were performed in the presence or absence of A3G; infectivity in the absence of A3G was normalized to 100%. Infectivity values (in percentage) are the average of six infections; error bars indicate the SD from the mean of these replicates. Infectivity of HIV-1ΔVif (light gray) and HIV-1 expressing SIVrcmCM8081 Vif (green), SIVcpzTan3 Vif (purple), or chimeric Vifs were tested against RCM A3G (top) or chimpanzee A3G (bottom). Constructs are depicted on the left, with SIVrcm and SIVcpz Vif fragments in green and purple, respectively, and with the PLP motif and the region of Vpr or Vpx overlap shown ("Vpr over" or "Vpx over") for chimera1 and chimera2.
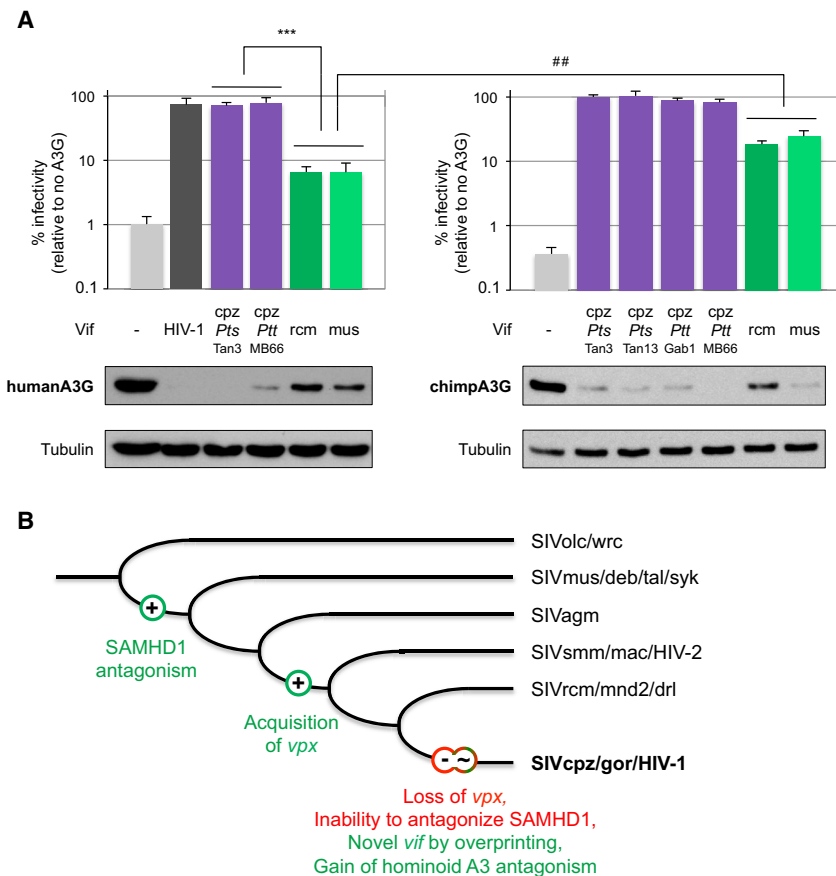
(legend continued on next page)

A

B

Figure 4. The Deep Origin of HIV-1 Lies in the Passage of OWM SIVs in the Chimpanzee Host and in the Evolution of *vif* and *vpx*

(A) Chimpanzee as a passage to human infection for OWM SIVs. Top panel: single-round infectivity assays were performed in the presence or absence of A3G as described in Figure 3. The graphs show the result for the average of six infections; error bars indicate the SD from the mean of these replicates. The infectivity of HIV-1ΔVif (gray, negative control) and HIV-1 plasmid with inserted HIV-1 LAI Vif (black); SIVcpz*Pts*Tan3, SIVcpz*Pts*Tan13, SIVcpz*Ptt*MB66, and SIVcpz*Ptt*Gab1 Vifs (purple); and SIVrcmCM8081 and SIVmus1CM1085 Vifs (green) was tested against human (left) or chimpanzee (right) A3Gs. HIV-1 and SIVcpz Vifs serve as positive controls against human and chimpanzee A3G, respectively. All Vifs were expressed at a level sufficient for their anti-A3G activity. ***p < 0.0001, ##p < 0.005, statistical difference by Mann-Whitney test. Lower panel: western blot analyses against HA-tagged A3G (left: human; right: chimpanzee) were performed on 293T cells cotransfected with the corresponding Vif constructs. Tubulin was probed as a loading control.

(B) The deep origin of HIV-1 is associated with the evolution of the accessory genes *vpx* and *vif* as a result of host antiviral gene-selective pressure. Shown is a representation of primate lentiviral evolution, highlighting major events that ultimately played a role in the origin of HIV-1. The main genetic events associated with *vpx* and *vif* are depicted along with SAMHD1 and A3G antagonism gain and loss. Green and positive signs are associated with a gain of function or gene acquisition/evolution, and red and negative signs are associated with a loss of function or gene loss.

with various Vif genes, using TransIT-LT1. The virus supernatants and the cells were harvested 48 hr posttransfection. Each transfection condition was performed in duplicate in independent experiments. The total amount of virus in the supernatant was quantified by p24 Gag ELISA (Advanced Bioscience Laboratories). SupT1 cells plated at 0.4 M cells/ml in the presence of 20 μg/ml diethylaminoethyl-dextran were infected with 2 ng of virus. Infections were performed in triplicate for 72 hr and luciferase activity was measured with the Bright-Glo Luciferase Assay Reagent (Promega). Statistical analyses were performed with the Mann-Whitney test. The harvested cells were used for western blot analyses.

Molecular Sequence Analyses
Alignments were performed using Muscle (Edgar, 2004) or FSA (Bradley et al., 2009), and minor adjustments were done when necessary. Maximum-likelihood trees were constructed using PhyML (Guindon et al., 2010) with 1,000 bootstrap replicates and the GTR model with four gamma rate categories. Recombination analyses were performed using GARD from Datamonkey (Kosakovsky Pond et al., 2006) and the PHI test (Bruen et al., 2006) implemented in SplitsTree (Huson and Bryant, 2006). The p value cutoff for any

evidence of recombination was set at 0.1. We used BLAST (Altschul et al., 1990), hmmer amino acid HMM search, and nhmmer (Eddy, 2011; Finn et al., 2011) to search for the origin of the 60–75 bp fragment in *vif* that is only present in the SIVcpz/HIV-1 lineage, first by looking in all sequences available in the NCBI database, and then more deeply by restricting the analyses to primate and SIV genomes (excluding this short fragment of the SIVcpz/HIV-1 lineage).

(C) SIVrcm Vif does not degrade chimpanzee A3G. Western blot analysis against HA-tagged A3G (left: RCM; right: chimpanzee) was performed on 293T cells cotransfected with HIV-1ΔVif, and HIV-1 expressing SIVcpzTan3Vif and SIVrcmCM8081Vif. Tubulin was probed as a loading control.
(D) Vif expression. Western blot analyses against FLAG-tagged Vifs from the HIV-1ΔVifΔEnvLuc backbone with the corresponding inserted Vif (SIV strains and chimeras correspond to B). Similar Vif expression relative to tubulin is shown.
(E) SIVrcm was pre-equipped to antagonize chimpanzee A3F and A3H, but not A3D. Single-round infectivity assays were performed against various chimpanzee APOBEC3 family proteins, as described in (B). Left: chimpanzee A3F; middle: A3H; right: A3D. Infectivity values (in percentage) are the average of six infections; error bars indicate the SD from the mean of these replicates. See also Figure S3.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ayinde, D., Maudet, C., Transy, C., and Margottin-Goguet, F. (2010). Limelight on two HIV/SIV accessory proteins in macrophage infection: is Vpx overshadowing Vpr? Retrovirology *7*, 35.

Bailes, E., Gao, F., Bibollet-Ruche, F., Courgnaud, V., Peeters, M., Marx, P.A., Hahn, B.H., and Sharp, P.M. (2003). Hybrid origin of SIV in chimpanzees. Science *300*, 1713.

Baldauf, H.M., Pan, X., Erikson, E., Schmidt, S., Daddacha, W., Burggraf, M., Schenkova, K., Ambiel, I., Wabnitz, G., Gramberg, T., et al. (2012). SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. Nat. Med. *18*, 1682–1687.

Barraud, P., Paillart, J.C., Marquet, R., and Tisné, C. (2008). Advances in the structural understanding of Vif proteins. Curr. HIV Res. *6*, 91–99.

Bergeron, J.R., Huthoff, H., Veselkov, D.A., Beavil, R.L., Simpson, P.J., Matthews, S.J., Malim, M.H., and Sanderson, M.R. (2010). The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS Pathog. *6*, e1000925.

Bishop, K.N., Verma, M., Kim, E.Y., Wolinsky, S.M., and Malim, M.H. (2008). APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Pathog. *4*, e1000231.

Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput. Biol. *5*, e1000392.

Bruen, T.C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. Genetics *172*, 2665–2681.

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

Donahue, J.P., Vetter, M.L., Mukhtar, N.A., and D'Aquila, R.T. (2008). The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation. Virology *377*, 49–53.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat. Rev. Immunol. *12*, 687–695.

Duggal, N.K., Malik, H.S., and Emerman, M. (2011). The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. J. Virol. *85*, 11361–11371.

Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*(Web Server issue), W29–W37.

Gaddis, N.C., Sheehy, A.M., Ahmad, K.M., Swanson, C.M., Bishop, K.N., Beer, B.E., Marx, P.A., Gao, F., Bibollet-Ruche, F., Hahn, B.H., and Malim, M.H. (2004). Further investigation of simian immunodeficiency virus Vif function in human cells. J. Virol. *78*, 12041–12046.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M., and Hahn, B.H. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature *358*, 495–499.

Gaur, R., and Strebel, K. (2012). Insights into the dual activity of SIVmac239 Vif against human and African green monkey APOBEC3G. PLoS ONE *7*, e48850.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., Florens, L., Washburn, M.P., and Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature *474*, 658–661.

Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., et al. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science *313*, 523–526.

Keese, P.K., and Gibbs, A. (1992). Origins of genes: "big bang" or continuous creation? Proc. Natl. Acad. Sci. USA *89*, 9489–9493.

Kirchhoff, F. (2010). Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. Cell Host Microbe *8*, 55–67.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D. (2006). GARD: a genetic algorithm for recombination detection. Bioinformatics *22*, 3096–3098.

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségéral, E., Yatim, A., Emiliani, S., Schwartz, O., and Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature *474*, 654–657.

Lim, E.S., Fregoso, O.I., McCoy, C.O., Matsen, F.A., Malik, H.S., and Emerman, M. (2012). The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe *11*, 194–204.

Malim, M.H., and Bieniasz, P.D. (2012). HIV restriction factors and mechanisms of evasion. Cold Spring Harb. Perspect. Med. *2*, a006940.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99–103.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J.F., et al. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J. Virol. *79*, 12515–12527.

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

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol. *2*, E275.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. *1*, a006841.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646–650.

Walker, R.C., Jr., Khan, M.A., Kao, S., Goila-Gaur, R., Miyagi, E., and Strebel, K. (2010). Identification of dominant negative human immunodeficiency virus type 1 Vif mutants that interfere with the functional inactivation of APOBEC3G by virus-encoded Vif. J. Virol. *84*, 5201–5211.

Zhang, C., de Silva, S., Wang, J.H., and Wu, L. (2012). Co-evolution of primate SAMHD1 and lentivirus Vpx leads to the loss of the vpx gene in HIV-1 ancestor. PLoS ONE *7*, e37477.