The evolution of HIV-1 and the origin of AIDS

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2010) **365**, 2487–2494
doi:10.1098/rstb.2010.0031

Review

The evolution of HIV-1 and the origin of AIDS

Paul M. Sharp[1,*] **and Beatrice H. Hahn**[2,3]

[1]*Institute of Evolutionary Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JT, UK*
[2]*Department of Medicine, and* [3]*Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA*

The major cause of acquired immune deficiency syndrome (AIDS) is human immunodeficiency virus type 1 (HIV-1). We have been using evolutionary comparisons to trace (i) the origin(s) of HIV-1 and (ii) the origin(s) of AIDS. The closest relatives of HIV-1 are simian immunodeficiency viruses (SIVs) infecting wild-living chimpanzees (*Pan troglodytes troglodytes*) and gorillas (*Gorilla gorilla gorilla*) in west central Africa. Phylogenetic analyses have revealed the origins of HIV-1: chimpanzees were the original hosts of this clade of viruses; four lineages of HIV-1 have arisen by independent cross-species transmissions to humans and one or two of those transmissions may have been via gorillas. However, SIVs are primarily monkey viruses: more than 40 species of African monkeys are infected with their own, species-specific, SIV and in at least some host species, the infection seems non-pathogenic. Chimpanzees acquired from monkeys two distinct forms of SIVs that recombined to produce a virus with a unique genome structure. We have found that SIV infection causes $CD4^+$ T-cell depletion and increases mortality in wild chimpanzees, and so the origin of AIDS is more ancient than the origin of HIV-1. Tracing the genetic changes that occurred as monkey viruses adapted to infect first chimpanzees and then humans may provide insights into the causes of the pathogenicity of these viruses.

Keywords: human immunodeficiency virus type 1; simian immunodeficiency virus; chimpanzee; gorilla; adaptation; tetherin

1. INTRODUCTION

The acquired immune deficiency syndrome (AIDS) was first formally recognized in patients in the USA in 1981. Subsequent characterization of the principal causative agent, human immunodeficiency virus type 1 (HIV-1), revealed that it was a retrovirus. As strains of HIV-1 were sampled from around the world, it became apparent that they exhibit extremely high genetic heterogeneity and that analysis of the evolution of this diversity can reveal insights into the prehistory of the virus (Sharp *et al.* 2001). HIV-1 strains can be divided into three distinct groups, which have very different prevalences. Groups N and O are rare, and largely restricted to Cameroon and surrounding countries. The vast majority (perhaps 98%) of HIV infections worldwide are caused by HIV-1 group M. Even within group M, there is very high diversity and the epicentre of that diversity is in Africa and in particular Kinshasa in the Democratic Republic of Congo (Vidal *et al.* 2000). While HIV-1 has an extremely fast rate of evolution, the virus must have circulated within human populations for many years before it was first recognized for this extent of diversity to have accumulated. Using molecular clocks, the common ancestor of HIV-1 group M strains has been dated to around the 1920s (Korber *et al.* 2000; Worobey *et al.* 2008). Partial characterization (Zhu *et al.* 1998; Worobey *et al.* 2008) of two viruses from samples initially obtained around 1960 in Kinshasa (then called Leopoldville) has shown that HIV-1 group M had already diversified substantially by that time, corroborating this time scale, and pointing to the location of the early diversification of these viruses (Sharp & Hahn 2008).

At the time when HIV-1 was first described, the closest known relative was visna, a virus from sheep that is the prototypic member of the genus Lentivirus. Additional lentiviruses were soon found in other primates, and a second virus (HIV-2) was found infecting humans. The viruses from non-human primates were termed simian immunodeficiency viruses (SIVs). Among the first species to be found to be naturally infected were African green monkeys (*Chlorocebus* species), where the prevalence of infection is high (greater than 50% of adults) and natural infections appear to be non-pathogenic. The number of different SIVs identified has increased steadily over the past 20 years. Currently, around 40 different primate species have been found to harbour SIVs, though information regarding prevalence and pathogenicity is lacking for most. So far, SIVs have only been found naturally infecting primates in sub-Saharan Africa, though the extent to which Asian or new world primates have been surveyed is unclear. Where multiple strains of SIVs have been characterized from a single species, they generally form a monophyletic clade, indicating that the great majority of

* Author for correspondence (paul.sharp@ed.ac.uk).

One contribution of 18 to a Discussion Meeting Issue 'Genetics and the causes of evolution: 150 years of progress since Darwin'.

transmissions are intraspecific. The primate viruses as a whole, including HIV-1 and HIV-2, form a distinct clade within the lentiviruses, indicating that humans acquired their infections from other primates (Bailes *et al.* 2002). Phylogenetic analyses of these primate lentiviruses have provided remarkably detailed insights into the evolutionary origins of the human viruses.

2. THE ORIGINS OF HIV-1 AND HIV-2

The origin of HIV-2 was resolved first. HIV-2 was first found, and is still most common, among individuals from west Africa. In 1989, a closely related SIV was found in a monkey, the sooty mangabey (*Cercocebus atys*), whose natural range is in west Africa (Hirsch *et al.* 1989). Other examples of this virus, termed SIVsmm, were soon found in other captive sooty mangabeys, and then in individuals from the wild (Chen *et al.* 1996; Santiago *et al.* 2005). Closely related viruses found in captive macaques cause severe AIDS-like illness (Letvin *et al.* 1985), but SIV has not been found in wild macaques (which are Asian primates), and SIVsmm seems to be non-pathogenic in its natural host (Santiago *et al.* 2005; Keele *et al.* 2009). It soon became apparent that SIVsmm had been inadvertently transmitted to various macaque species in captivity (reviewed in Apetrei *et al.* 2005). In phylogenetic analyses, HIV-2 strains can be divided into various groups, which are interspersed among the SIVsmm lineages (Gao *et al.* 1992, 1994). These observations led to a rapid acceptance that sooty mangabeys were the source of HIV-2, and the interspersion of HIV-2 strains among the SIVsmm lineages implied that there had been multiple mangabey-to-human cross-species jumps (Hahn *et al.* 2000). Only two of these transmission events, leading to groups A and B, have given rise to viruses that have spread widely in the human population; six other lineages are each known only from viruses found in single individuals (Damond *et al.* 2004).

A virus closely related to HIV-1 was also first reported in 1989; this virus, SIVcpz, was found in two captive chimpanzees (*Pan troglodytes*) in Gabon (Peeters *et al.* 1989; Huet *et al.* 1990). However, for many years, chimpanzees were not accepted as the source of HIV-1 because it remained unclear whether wild chimpanzees are naturally infected with this virus. Over the next 10 years, many (probably more than a thousand) other chimpanzees were tested but only a single further example of SIVcpz was found, in a chimpanzee illegally imported to Belgium from Kinshasa (Peeters *et al.* 1992). Furthermore, this third example of SIVcpz was more divergent from the previous examples than might be expected for two viruses from a single host species (Vanden Haesevelde *et al.* 1996). Thus, it appeared that SIV was extremely rare in chimpanzees, and it seemed possible that both chimpanzees and humans had been infected from some other source(s), presumably some monkey species in central Africa.

In 1999, we found a fourth example of SIV from a chimpanzee. This ape was wild-caught and imported to the USA, but records of its geographical origin were not available. Chimpanzees have traditionally been divided into several subspecies (Groves 2001). Analyses of mitochondrial DNA (mtDNA) indicate four subspecies: western (*Pan troglodytes verus*), Nigerian (*Pan t. ellioti*), central (*P. t. troglodytes*) and eastern (*Pan t. schweinfurthii*) chimpanzees (Gagneux *et al.* 1999), which have non-overlapping ranges across western and central Africa (figure 1). (Note that *P. t. ellioti* was formerly termed *P. t. vellerosus*; Oates *et al.* 2009.) From mtDNA sequence analyses, we found that the two chimpanzees from Gabon were (as expected) *P. t. troglodytes*, as was the ape from the USA, while the other individual belonged to *P. t. schweinfurthii* (Gao *et al.* 1999). In contrast, it seemed likely that the vast majority of chimpanzees that had tested negative for SIVs were *P. t. verus*; this was subsequently confirmed (Prince *et al.* 2002; Switzer *et al.* 2005). Thus, the apparent scarcity of SIVs in chimpanzees could be explained by an absence of infection in the one subspecies that had been subject to the most testing. Indeed, two additional examples of SIVcpz from *P. t. troglodytes* were soon reported (Corbet *et al.* 2000).

The SIVcpz strains from *P. t. troglodytes* individuals formed a monophyletic cluster with all HIV-1 strains, with the single SIV from a *P. t. schweinfurthii* ape forming an outgroup. All subsequently characterized SIVcpz strains have conformed to this host subspecies-specific clustering, and this explained the unusually high level of divergence between the isolates from Gabon and Belgium. These observations implicated one particular chimpanzee subspecies, *P. t. troglodytes*, as the source of the human viruses (Gao *et al.* 1999). Since the various HIV-1 groups are interspersed among the SIVcpz lineages, each must have had a separate origin (figure 2).

Before 2002, all SIVs were isolated from blood or tissue samples. This severely limited the extent to which rare and endangered primate species could be tested for more SIV strains, and in particular prevented estimation of the prevalence of SIVs in wild chimpanzee populations. Therefore, we developed an approach for non-invasive screening of wild primates, using faecal samples; from these, we are able to detect antibodies against SIVs and obtain nucleotide sequences from both the host and the virus (Santiago *et al.* 2002). The species and subspecies, as well as the sex and identity, of the host can be established from mtDNA and various nuclear DNA markers, respectively. Using this non-invasive sampling technique, we have been able to show that SIVcpz infection is quite common and widespread among both central and eastern chimpanzees (Santiago *et al.* 2003; Worobey *et al.* 2004; Keele *et al.* 2006, 2009; Van Heuverswyn *et al.* 2007). However, the prevalence is patchy: in some areas, around one-third of individuals are infected, whereas in other areas the virus appears to be absent. These surveys have finally established that chimpanzees are indeed the natural reservoir for SIVcpz and the source of HIV-1, and have also supported the contention that *P. t. verus*, and in addition *P. t. ellioti*, are not infected with SIVs.

Initial phylogenetic analyses of the viral sequences obtained by non-invasive sampling revealed two key points (Keele *et al.* 2006). First, among the

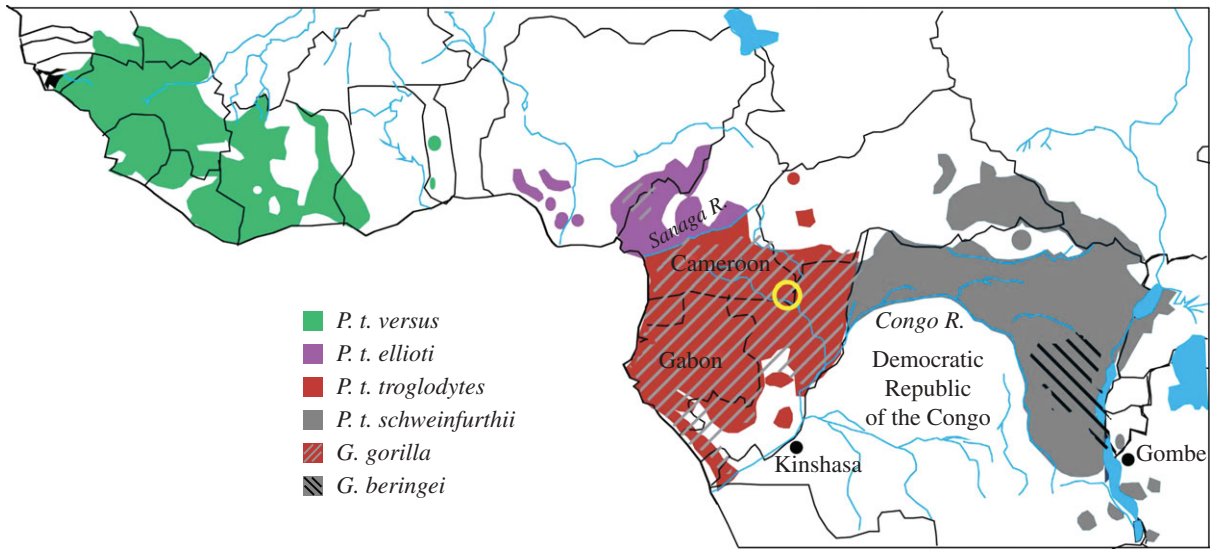

Figure 1. Map of west and central Africa, showing the ranges of chimpanzee subspecies (colour coded). The ranges of western lowland gorillas (*G. gorilla*) and eastern Grauer's gorillas (*G. beringei*) are superimposed. The gold circle denotes the region in southeast Cameroon where SIVcpz strains closely related to HIV-1 group M are found.

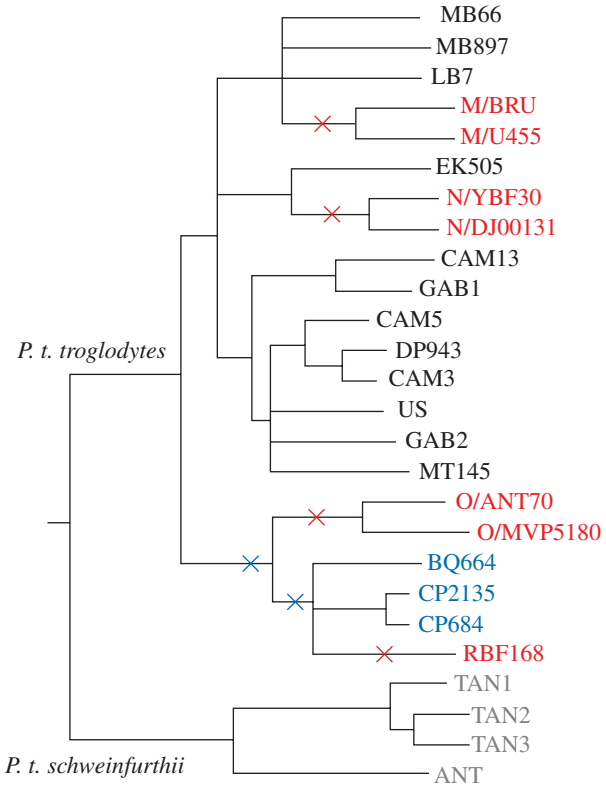

Figure 2. Origins of HIV-1. The phylogenetic relationships among strains of SIVcpz (black from *P. t. troglodytes*, grey from *P. t. schweinfurthii*), SIVgor (blue) and HIV-1 (red). The red crosses mark four branches on which cross-species jumps to humans occurred; the two blue crosses indicate alternative possible branches on which a chimpanzee-to-gorilla transmission occurred. The HIV-1 strains fall into three groups (M, N and O; only two representatives of each group are shown), and a recently described fourth lineage (RBF168). Adapted from trees shown in Takehisa *et al.* (2009) and Plantier *et al.* (2009).

newly described viruses were two clades of strains extremely closely related to HIV-1 groups M and N, respectively (figure 2). Second, the strains of SIVcpz displayed phylogeographic clustering; that is, multiple strains obtained from any one location formed monophyletic clusters, while sample sites separated by geographical barriers (typically large rivers) either were discordant with regard to the presence or absence of SIV infections, or featured strains of SIVs that were not especially closely related. This observation strongly suggests that the transmissions from chimpanzee to human that gave rise to groups M and N most probably occurred in the locations where the closely related SIVcpz strains were found. In particular, these analyses pinpoint the probable source of the viruses that gave rise to the HIV-1 group M pandemic as being chimpanzees in the extreme southeast corner of Cameroon (figure 1), in an area flanked by the Boumba, Ngoko and Sangha rivers (Keele *et al.* 2006).

While HIV-1 group O also falls within the radiation of SIVcpz strains, none of the chimpanzee viruses were particularly closely related to HIV-1 group O. Further analyses of ape faecal samples from across Cameroon revealed group O-related viruses in gorillas (*Gorilla gorilla gorilla*) (Van Heuverswyn *et al.* 2006; Takehisa *et al.* 2009). Very recently, Plantier *et al.* (2009) have reported an HIV-1 strain that does not fall within any of the three described groups, and so must represent a fourth cross-species transmission; if viruses from this lineage are found in other individuals, they will constitute a fourth group, P. This new strain of HIV-1 is very closely related to the gorilla viruses (SIVgor), and has most probably resulted from gorilla-to-human transmission. At present, it is unclear whether gorillas were also the immediate source of HIV-1 group O, or whether chimpanzee viruses were transmitted in parallel to gorillas and humans (figure 2).

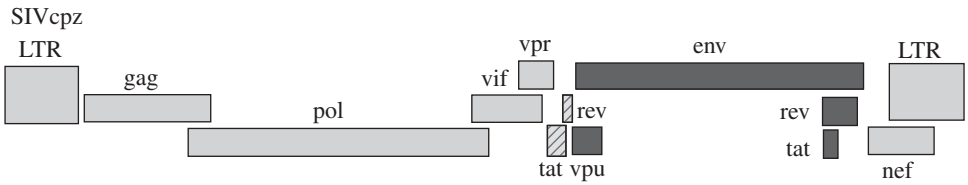

Figure 3. Mosaic structure of the SIVcpz genome. Boxes denote the long terminal repeat (LTR) regions and nine genes (*tat* and *rev* each have two exons). The SIVcpz genome arose through recombination; regions in light and dark grey were derived from the SIVrcm and SIVgsn/SIVmus/SIVmon lineages, respectively (the origin of the 5′ exons of *tat* and *rev* is difficult to determine).

3. THE ORIGIN OF AIDS

The conclusion that HIV-1 was derived from a virus infecting chimpanzees is particularly interesting, because chimpanzees and humans are so closely related. This raises a number of questions, such as the origin of the chimpanzee virus, whether adaptation of SIVcpz to infecting chimpanzees rendered the virus more capable of infecting humans and whether SIVcpz infection of chimpanzees is pathogenic.

SIVagm infections of African green monkeys and SIVsmm infections of sooty mangabeys appear to be non-pathogenic (Silvestri 2008). Extrapolating from this, it has been generally assumed that all natural SIV infections, including that of chimpanzees, are harmless. If true, the origin of AIDS coincided with the origin of HIV. However, we have recently reported evidence that contradicts this (Keele *et al*. 2009). The SIV infection status of two habituated communities of chimpanzees at Gombe National Park in Tanzania has been studied since 2001 (Santiago *et al*. 2002). From more than 550 chimpanzee-years of observations, we found that SIV infection was associated with a 10- to 16-fold increase in age-corrected risk of death. It was also found that fertility was significantly reduced in SIV-positive females, both in terms of their birth rate and the survival of their offspring. The primary symptom of AIDS in humans is a reduction in the number of $CD4^+$ T-cells; the depletion of these cells reduces host defences against secondary infections. It was possible to determine $CD4^+$ T-cell counts for five deceased chimpanzees. Tissues from three SIV-positive chimpanzees had significantly lower counts than those from two SIV-negative individuals. Two of the three SIV-positive individuals had died of trauma-related causes, while the third had no obvious injuries but displayed weakness and lethargy; this third chimpanzee had the lowest $CD4^+$ T-cell counts and tissue samples from this individual closely resembled the histopathology of human end-stage AIDS patients (Keele *et al*. 2009). While it has only so far been possible to assess the pathogenicity of SIVcpz at Gombe, where the apes are eastern chimpanzees (figure 1), there is no reason to believe that SIVcpz infection of central chimpanzees differs in any substantial way. These observations strongly suggest that SIVcpz infections in wild chimpanzees have a very similar effect to HIV-1 infections of humans.

Since most SIVs infect monkey species, whereas chimpanzees (like humans) are apes, it has always seemed likely that chimpanzees initially acquired SIVs from monkeys. However, none of the monkey SIVs described prior to 2001 were especially closely related to SIVcpz. When SIVrcm was found in red-capped mangabeys (*Cercocebus torquatus*), it was noted that the 5′ half of the genome was most similar to SIVcpz (Beer *et al*. 2001). Soon after, a virus (SIVgsn) was found in greater spot-nosed monkeys (*Cercopithecus nictitans*), where the 3′ half of the genome was most similar to SIVcpz (Courgnaud *et al*. 2002). Both SIVrcm and SIVgsn were interpreted as being recombinants between SIVcpz and other, as yet unidentified, SIV lineages. However, we showed that it is more likely that SIVcpz arose from recombination between the SIVrcm and SIVgsn lineages (Bailes *et al*. 2003), obviating the need to invoke other unknown SIVs. Subsequently, viruses (SIVmus and SIVmon) closely related to SIVgsn were found in mustached guenons (*Cercopithecus cephus*) and mona monkeys (*Cercopithecus mona*) (Courgnaud *et al*. 2003), and it can be seen that it was an ancestor of these three SIVs that was involved in the recombination event that generated SIVcpz (Sharp *et al*. 2005). Thus, it appears that chimpanzees acquired SIVs from two different species of monkeys, most probably by predation. The current ranges of red-capped mangabeys and the relevant *Cercopithecus* species overlap that of *P. t. troglodytes* in west central Africa (Groves 2001), and so it appears likely that this subspecies was first infected. The two monkey viruses then recombined, and this mosaic spread to become the only form now found in chimpanzees, and the ancestor of HIV-1 and SIVgor.

This recombinant virus has a genome structure unique among the SIVs (figure 3). First, it has a *vpu* gene overlapping the 5′ end of the *env* gene. This *vpu* gene is not present in the genomes of most monkey SIVs (including SIVrcm), but is found in viruses from the lineage including SIVgsn, SIVmus and SIVmon. Second, while in most SIV genomes there is a short overlap between the 3′ end of the *env* gene and the 5′ end of the *nef* gene, in SIVcpz there is none. This region has been duplicated in SIVcpz, presumably during the recombination event, because each of the duplicate copies was derived from a different parental virus (Schindler *et al*. 2006).

It has been found that most SIVs, and in particular those for which there is the best evidence of non-pathogenicity in the natural host (SIVsmm and SIVagm), encode a Nef protein that downregulates the T-cell receptor CD3 and thereby suppresses T-cell activation; in contrast, the Nef protein of SIVcpz, like that of HIV-1, does not downregulate CD3 (Schindler *et al*. 2006). This is surprising because the *nef* gene of SIVcpz was derived from the SIVrcm

lineage (figure 3) and SIVrcm Nef does downregulate CD3 (Schindler *et al.* 2006); thus, the Nef protein of SIVcpz has lost this activity. Since high-level immune activation is associated with progression to AIDS, this change in the properties of SIVcpz Nef correlates with pathogenicity. It is not known whether SIVgsn, SIVmus or SIVmon are pathogenic in their natural hosts. The Nef proteins of these viruses also do not downregulate CD3, but since the *nef* gene of SIVcpz was not acquired from the SIVgsn lineage, this property must have evolved independently in SIVcpz. The shared feature of the viruses that do not downregulate CD3 is the presence of a *vpu* gene, but it is not clear why this might have prompted the *nef* gene to evolve to lose this function.

4. HOST ADAPTATION

HIVs and SIVs interact with many host proteins (Bushman *et al.* 2009). Since many of those host proteins have diverged since the common ancestor of Old World monkeys and apes, it is likely that when chimpanzees first acquired SIV, the virus had to adapt in order to replicate efficiently and spread in its new host. Chimpanzees and humans are genetically very similar owing to their recent common ancestry, but there is also evidence that differences between chimpanzee and human proteins placed selection pressures on SIVcpz after it jumped from chimpanzees to humans.

Tetherin is a mammalian protein with a recently discovered antiviral activity. Tetherin dimers appear to form 'tethers' between virus envelopes and the cytoplasmic membrane of the cell, preventing the release of those viruses (Neil *et al.* 2008). The Vpu protein of HIV-1 has long been known to promote the release of progeny virions, and it is now known that this is mediated by antagonizing tetherin. However, as described above, the *vpu* gene is present in only a few SIV genomes. In SIVsmm and SIVagm, which lack *vpu* genes, the Nef protein has anti-tetherin activity (Jia *et al.* 2009; Zhang *et al.* 2009). Nevertheless, it was surprising to find that in SIVcpz, it is the Nef protein, rather than the Vpu protein, that counteracts tetherin (Sauter *et al.* 2009). When SIVcpz was formed by recombination, it acquired two genes with potential anti-tetherin activity: a *vpu* gene from the SIVgsn lineage and a *nef* gene from SIVrcm (figure 3); the Vpu protein of SIVgsn has been shown to counteract greater spot-nosed monkey tetherin (Sauter *et al.* 2009), while the Nef protein of SIVrcm is expected to function against mangabey tetherin (although this has not yet been tested experimentally). However, there is species specificity in these interactions, reflecting divergence in the tetherin sequences of different primates. Thus, when SIVs first infected chimpanzees, it is likely that neither the Vpu nor the Nef protein had full anti-tetherin activity; certainly Vpu proteins from contemporary SIVgsn, SIVmon or SIVmus strains have no activity against chimpanzee tetherin (Sauter *et al.* 2009). Vpu and Nef antagonize tetherin through interactions with different parts of the tetherin protein: for Vpu it is the transmembrane (TM) region, and for Nef the cytoplasmic tail (CT). In a comparison of monkey and chimpanzee tetherin sequences, the CT is less divergent than the TM and so this may be why the Nef protein was better able to adapt.

Since SIVcpz uses the Nef protein to counteract tetherin, it is surprising that HIV-1 uses Vpu. However, human tetherin has diverged from the chimpanzee protein, most notably owing to a deletion of a pentamer within the CT. As a consequence, the SIVcpz Nef protein is not active against human tetherin (Sauter *et al.* 2009). Clearly, this would have placed strong selection pressure on HIV-1, and this has resulted in the reacquisition of an anti-tetherin activity by Vpu. Perhaps the surprising aspect of this is that, while the Vpu protein of SIVcpz was not being constrained to retain anti-tetherin activity, nevertheless the *vpu* gene did not diverge to the extent that the activity could not be rescued.

Since the HIV-1 groups M, N and O each arose through separate transmissions of SIVs from apes, the selection pressure to counteract human tetherin would have been exerted on three independent occasions. This led to different outcomes. Only in the case of HIV-1 group M has adaptation been fully successful. In HIV-1 group O strains, neither Vpu nor Nef efficiently antagonize human tetherin (Sauter *et al.* 2009). HIV-1 group N Vpu has evolved to become active against tetherin, but this appears to have been at a cost. A second major function of Vpu is to bind to CD4 to induce its degradation. Unlike the anti-tetherin activity, this function seems to be highly conserved and not species-specific, perhaps because CD4 itself is less divergent among primates than is tetherin. However, HIV-1 group N Vpu has lost this anti-CD4 activity, possibly in the process of adaptation to an anti-tetherin activity. In summary, HIV-1 group M Vpu has two main functions: an anti-CD4 activity that was conserved from SIVcpz, and an anti-tetherin activity acquired since transmission to humans. HIV-1 group N has lost the first of these, while HIV-1 group O has not acquired the second. It is tempting to speculate that these differences may at least partly explain why group N and O viruses are quite rare, while group M viruses dominate the global AIDS pandemic (Sauter *et al.* 2009).

An alternative approach to seeking signatures of host-specific adaptation in HIV-1 is to look for similar genetic changes that occurred independently on each of the occasions when SIVs jumped into humans. Comparing the sequences of SIVcpz with those of the inferred ancestors of the three HIV-1 groups, we found one site in the proteome that was well conserved among SIVcpz but had changed, in the same way, on each of the three host jumps giving rise to groups M, N and O (Wain *et al.* 2007). Codon 30 of the *gag* gene encodes Met in all known strains of SIVcpz from central chimpanzees, and also in SIVgor. However, the inferred ancestral sequences of the three HIV-1 groups encode Arg at this position, implying that a radical amino acid replacement occurred on each of the three branches of the tree encompassing cross-species transmission to humans. Two lines of evidence provide corroboration that this site is involved in host-species-specific adaptation.

The first comes from an experiment in which a chimpanzee was infected with HIV-1; when the virus was sequenced 10 years later, this site had reverted to encoding Met (Mwaengo & Novembre 1998). Secondly, this chimpanzee-adapted HIV-1 virus was subjected to site-directed mutagenesis and then tested for replication in chimpanzee and human $CD4^+$ T-cells. Viruses differing only at codon 30 of *gag* grew with different efficiency in chimpanzee T-cells, with those encoding Met replicating faster; the opposite was observed in human T-cells (Wain *et al*. 2007). This part of the *gag* gene encodes the N-terminal domain of the matrix protein. The structure of this protein has been solved (Hill *et al*. 1996), and while there is considerable information about the interaction of the matrix protein with host proteins, there is as yet no clue as to the function of this particular residue or why chimpanzee and human hosts exert different selection pressures on it.

5. CONCLUSIONS

It has been possible to reconstruct a surprisingly detailed picture of the origins of pandemic AIDS (Sharp & Hahn 2008). The source of HIV-1 group M, the main form of AIDS virus infecting humans, has been traced to a virus infecting the central subspecies of chimpanzees, *P. t. troglodytes*, in a remote area in the southeast corner of Cameroon (Keele *et al*. 2006). The likeliest route of chimpanzee-to-human transmission would have been through exposure to infected blood and body fluids during the butchery of bushmeat (Hahn *et al*. 2000). The early diversification of group M appears to have occurred some 700 km further south, in Kinshasa (then called Leopoldville), in the early years of the twentieth century (Worobey *et al*. 2008). The links between these two regions are waterways, which were the major communication routes at that time: the rivers from southeast Cameroon flow south, ultimately joining the Congo River on which Kinshasa is located. Other viruses from central chimpanzees have been transmitted to humans, giving rise to HIV-1 groups N and O, and the prospective group P; among these, some (probably P, and perhaps O) may have been transmitted via gorillas. It is striking that all of these transmissions were derived from one subspecies of chimpanzee. Eastern chimpanzees also have SIVcpz, but descendants of those viruses have not been detected in humans, nor in gorillas that are sympatric with eastern chimpanzees (Takehisa *et al*. 2009). It is not clear that humans living in east central Africa are less likely to have been exposed to SIVcpz than those living in west central Africa. It is possible that the level of surveillance has been lower in east central Africa, and that rare strains of HIV-1 derived from the SIVcpz of eastern chimpanzees circulate undetected. Alternatively, given that it is clear that SIVcpz of central chimpanzees has had to adapt in various ways to spread in humans, it is possible that the divergent form of SIVcpz found in eastern chimpanzees has less potential to make these adaptations.

Considering HIV-1 and HIV-2 together, it appears that there have been at least a dozen independent transmissions of SIVs to humans. None of these appear older than HIV-1 group M: the common ancestor of HIV-1 group O has been dated to around the same time (Lemey *et al*. 2004), while those of HIV-2 groups A and B may have been a little more recent (Lemey *et al*. 2003). The opportunities for chimpanzee- or monkey-to-human host jumps have existed for hundreds or thousands of years, and it must be expected that many such transmissions occurred in the past. However, only in the twentieth century did such viruses spread to detectable levels in the human population. In west central Africa during the early part of that century, the destabilization of social structures by invading colonial powers (Chitnis *et al*. 2000), the origin and rapid growth of major conurbations (Worobey *et al*. 2008) and the widespread use of injections (Pepin & Labbe 2008) may all have contributed to provide an unprecedented opportunity for dissemination of blood-borne viruses.

The finding that HIV-1 originated in our closest living relatives raises a number of issues. Gene sequences of humans and chimpanzees typically differ at less than 2 per cent of nucleotides. Despite this genetic similarity of humans and chimpanzees, it is apparent that SIVcpz was subject to pressures to adapt to its new human host. It is clearly of substantial interest to understand much more about the natural biology of SIVcpz infection and transmission in chimpanzees, and the extent to which the natural history of these viruses differs from HIV-1 in humans. It will be of particular interest to understand the extent to which co-infection with other viruses, bacteria, protozoa (such as *Plasmodium*) or multicellular eukaryotes (e.g. worms) influence the course of SIVcpz infection and pathogenesis. For example, there is some evidence that humans co-infected with HIV-1 and GBV-C (a seemingly non-pathogenic member of the Flaviviridae) remain asymptomatic longer than those infected by HIV-1 alone (Stapleton *et al*. 2004); close relatives of GBV-C have been found in chimpanzees (Adams *et al*. 1998; Birkenmeyer *et al*. 1998). Further comparisons of SIVcpz and HIV-1 infections may shed light on the viral and host factors responsible for disease progression and ultimately point the way to novel therapeutic interventions.

We thank Jamie White for assistance in producing the figures. This work was supported by grants from National Institutes of Health (R01 AI50529, R01 AI58715, P30 AI 27767, P30 CA13148) and the Bristol Myers Freedom to Discover programme.

REFERENCES

Adams, N. J., Prescott, L. E., Jarvis, L. M., Lewis, J. C. M., McClure, M. O., Smith, D. B. & Simmonds, P. 1998 Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *J. Gen. Virol.* **79**, 1871–1877.

Apetrei, C. *et al.* 2005 Molecular epidemiology of simian immunodeficiency virus SIVsm in US primate centers unravels the origin of SIVmac and SIVstm. *J. Virol.* **79**, 8991–9005. (doi:10.1128/JVI.79.14.8991-9005.2005)

Bailes, E., Chaudhuri, R. R., Santiago, M. L., Bibollet-Ruche, F., Hahn, B. H. & Sharp, P. M. 2002 The evolution of primate lentiviruses and the origins of

AIDS. In *The Molecular epidemiology of human viruses* (ed. T. Leitner), pp. 65–96. Boston, MA: Kluwer Academic Publishers.

Bailes, E., Gao, F., Bibollet-Ruche, F., Courgnaud, V., Peeters, M., Marx, P. A., Hahn, B. H. & Sharp, P. M. 2003 Hybrid origin of SIV in chimpanzees. *Science* **300**, 1713. (doi:10.1126/science.1080657)

Beer, B. E. *et al.* 2001 Characterization of novel simian immunodeficiency viruses from red-capped mangabeys from Nigeria (SIVrcmNG409 and –Ng411). *J. Virol.* **75**, 12 014–12 027. (doi:10.1128/JVI.75.24.12014-12027.2001)

Birkenmeyer, L. G., Desai, S. M., Muerhoff, A. S., Leary, T. P., Simons, J. N., Montes, C. C. & Mushahwar, I. K. 1998 Isolation of a GB virus-related genome from a chimpanzee. *J. Med. Virol.* **56**, 44–51. (doi:10.1002/(SICI)1096-9071(199809)56:1<44::AID-JMV8>3.0.CO;2-N)

Bushman, F. D. *et al.* 2009 Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* **5**, e1000437. (doi:10.1371/journal.ppat.1000437)

Chen, Z., Telfer, P., Gettie, A., Reed, P., Zhang, L., Ho, D. D. & Marx, P. A. 1996 Genetic characterization of new west African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol.* **70**, 3617–3627.

Chitnis, A., Rawls, D. & Moore, J. 2000 Origin of HIV type 1 in colonial French Equatorial Africa? *AIDS Res. Hum. Retroviruses* **16**, 5–8. (doi:10.1089/088922200309548)

Corbet, S. *et al.* 2000 *Env* sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* **74**, 529–534. (doi:10.1128/JVI.74.1.529-534.2000)

Courgnaud, V. *et al.* 2002 Characterization of a novel simian immunodeficiency virus with a *vpu* gene from greater spot-nosed monkeys (*Cercopithecus nictitans*) provides new insights into simian/human immunodeficiency virus phylogeny. *J. Virol.* **76**, 8298–8309. (doi:10.1128/JVI.76.16.8298-8309.2002)

Courgnaud, V., Abela, B., Pourrut, X., Mpoudi-Ngole, E., Loul, S., Delaporte, E. & Peeters, M. 2003 Identification of a new simian immunodeficiency virus lineage with a *vpu* gene present among different *Cercopithecus* monkeys (*C. mona*, *C. cephus*, and *C. nictitans*) from Cameroon. *J. Virol.* **77**, 12 523–12 534. (doi:10.1128/JVI.77.23.12523-12534.2003)

Damond, F., Worobey, M., Campa, P., Farfara, I., Colin, G., Matherson, S., Brin-Vezinet, F., Robertson, D. L. & Simon, F. 2004 Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res. Hum. Retroviruses* **20**, 666–672. (doi:10.1089/0889222041217392)

Gagneux, P. *et al.* 1999 Mitochondrial sequences show diverse evolutionary histories of African hominids. *Proc. Natl Acad. Sci. USA* **96**, 5077–5082. (doi:10.1073/pnas.96.9.5077)

Gao, F. *et al.* 1992 Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. *Nature* **358**, 495–499. (doi:10.1038/358495a0)

Gao, F. *et al.* 1994 Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* **68**, 7433–7447.

Gao, F. *et al.* 1999 Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436–441. (doi:10.1038/17130)

Groves, C. 2001 *Primate taxonomy.* Washington, DC: Smithsonian Institution Press.

Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. 2000 AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614. (doi:10.1126/science.287.5453.607)

Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M. & Sundquist, W. I. 1996 Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl Acad. Sci. USA* **93**, 3099–3104. (doi:10.1073/pnas.93.7.3099)

Hirsch, V. M., Olmsted, R. A., Murphy-Corb, M., Purcell, R. H. & Johnson, P. R. 1989 An African primate lentivirus (SIV_{sm}) closely related to HIV-2. *Nature* **339**, 389–392. (doi:10.1038/339389a0)

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G. & Wain-Hobson, S. 1990 Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356–358. (doi:10.1038/345356a0)

Jia, B., Serra-Moreno, R., Neidermyer, W., Rahmberg, A., Mackey, J., Fofana, I. B., Johnson, W. E., Westmoreland, S. & Evans, D. T. 2009 Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. *PLoS Pathog.* **5**, e1000429. (doi:10.1371/journal.ppat.1000429)

Keele, B. F. *et al.* 2006 Chimpanzee reservoirs of pandemic and non-pandemic HIV-1. *Science* **313**, 523–526. (doi:10.1126/science.1126531)

Keele, B. F. *et al.* 2009 Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* **460**, 515–519. (doi:10.1038/nature08200)

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. 2000 Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796. (doi:10.1126/science.288.5472.1789)

Lemey, P., Pybus, O. G., Wang, B., Saksena, N. K., Salemi, M. & Vandamme, M. 2003 Tracing the origin and history of the HIV-2 epidemic. *Proc. Natl Acad. Sci. USA* **100**, 6588–6592. (doi:10.1073/pnas.0936469100)

Lemey, P., Pybus, O. G., Rambaut, A., Drummond, A. J., Robertson, D. L., Roques, P., Worobey, M. & Vandamme, M. 2004 The molecular population genetics of HIV-1 group O. *Genetics* **167**, 1059–1068. (doi:10.1534/genetics.104.026666)

Letvin, N. L. *et al.* 1985 Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III. *Science* **230**, 71–73. (doi:10.1126/science.2412295)

Mwaengo, D. M. & Novembre, F. J. 1998 Molecular cloning and characterization of viruses isolated from chimpanzees with pathogenic human immunodeficiency virus infections. *J. Virol.* **72**, 8976–8987.

Neil, S. J. D., Zang, T. & Bieniasz, P. D. 2008 Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425–430. (doi:10.1038/nature06553)

Oates, J. F., Groves, C. P. & Jenkins, P. D. 2009 The type locality of *Pan troglodytes vellerosus* (Gray, 1862), and implications for the nomenclature of West African chimpanzees. *Primates* **50**, 78–80. (doi:10.1007/s10329-008-0116-z)

Peeters, M., Honore, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R. W. & Delaporte, E. 1989 Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* **3**, 625–630.

Peeters, M., Fransen, K., Delaporte, E., Van den Haesevelde, M., Gershy-Damet, G.-M., Kestens, L., van der Groen, G. & Piot, P. 1992 Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *AIDS* **3**, 625–630.

Pepin, J. & Labbe, C. 2008 Noble goals, unforeseen consequences: control of tropical diseases in colonial Central Africa and the iatrogenic transmission of blood-borne viruses. *Trop. Med. Int. Health* **13**, 744–753.

Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. & Simon, F. 2009 A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15**, 871–872. (doi:10.1038/nm.2016)

Prince, A. M., Brotman, B., Lee, D.-H., Andrus, L., Valinsky, J. & Marx, P. A. 2002 Lack of evidence for HIV type 1-related SIVcpz infection in captive and wild chimpanzees (*Pan troglodytes verus*) in West Africa. *AIDS Res. Hum. Retroviruses* **18**, 657–660. (doi:10.1089/088922202760019356)

Santiago, M. L. *et al.* 2002 SIVcpz in wild chimpanzees. *Science* **295**, 465. (doi:10.1126/science.295.5554.465)

Santiago, M. L. *et al.* 2003 Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J. Virol.* **77**, 7545–7562. (doi:10.1128/JVI.77.13.7545-7562.2003)

Santiago, M. L. *et al.* 2005 Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J. Virol.* **79**, 12 515–12 527. (doi:10.1128/JVI.79.19.12515-12527.2005)

Sauter, D. *et al.* 2009 Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and non-pandemic HIV-1 strains. *Cell Host Microbe* **6**, 409–421. (doi:10.1016/j.chom.2009.10.004)

Schindler, M. *et al.* 2006 Nef-mediated suppression of T cell activation is a fundamental property of primate lentiviruses that was lost in the lineage that gave rise to HIV-1. *Cell* **125**, 1055–1067. (doi:10.1016/j.cell.2006.04.033)

Sharp, P. M. & Hahn, B. H. 2008 Prehistory of HIV-1. *Nature* **455**, 605–606. (doi:10.1038/455605a)

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O. & Hahn, B. H. 2001 The origins of AIDS viruses: where and when? *Phil. Trans. R. Soc. Lond. B* **356**, 867–876. (doi:10.1098/rstb.2001.0863)

Sharp, P. M., Shaw, G. M. & Hahn, B. H. 2005 Simian immunodeficiency virus infection of chimpanzees. *J. Virol.* **79**, 3891–3902. (doi:10.1128/JVI.79.7.3891-3902.2005)

Silvestri, G. 2008 Immunity in natural SIV infections. *J. Intern. Med.* **265**, 97–109. (doi:10.1111/j.1365-2796.2008.02049.x)

Stapleton, J. T., Williams, C. F. & Xiang, J. 2004 GB virus type C: a beneficial infection. *J. Clin. Microbiol.* **42**, 3915–3919. (doi:10.1128/JCM.42.9.3915-3919.2004)

Switzer, W. M., Parekh, B., Shanmugam, V., Bhullar, V., Phillips, S., Ely, J. J. & Heneine, W. 2005 *AIDS Res. Hum. Retroviruses* **21**, 335–342. (doi:10.1089/aid.2005.21.335)

Takehisa, J. *et al.* 2009 Origin and biology of simian immunodeficiency virus infecting wild western gorillas. *J. Virol.* **83**, 1635–1648. (doi:10.1128/JVI.02311-08)

Vanden Haesevelde, M. M., Peeters, M., Jannes, G., Janssens, W., van der Groen, G., Sharp, P. M. & Saman, E. 1996 Sequence analysis of a highly divergent HIV-1 related lentivirus isolated from a wild captured chimpanzee. *Virology* **221**, 346–350. (doi:10.1006/viro.1996.0384)

Van Heuverswyn, F. *et al.* 2006 SIV infection in wild gorillas. *Nature* **444**, 164. (doi:10.1038/444164a)

Van Heuverswyn, F. *et al.* 2007 Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. *Virology* **368**, 155–171.

Vidal, N. *et al.* 2000 Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in central Africa. *J. Virol.* **74**, 10 498–10 507. (doi:10.1128/JVI.74.22.10498-10507.2000)

Wain, L. V. *et al.* 2007 Adaptation of HIV-1 to its human host. *Mol. Biol. Evol.* **24**, 1853–1860. (doi:10.1093/molbev/msm110)

Worobey, M. *et al.* 2004 Origin of AIDS: contaminated polio vaccine theory refuted. *Nature* **428**, 820. (doi:10.1038/428820a)

Worobey, M. *et al.* 2008 Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664. (doi:10.1038/nature07390)

Zhang, F. *et al.* 2009 Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell Host Microbe* **6**, 54–67. (doi:10.1016/j.chom.2009.05.008)

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M. & Ho, D. D. 1998 An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594–597.