THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0863

The origins of acquired immune deficiency syndrome viruses: where and when?

Paul M. Sharp[1*], Elizabeth Bailes[1], Roy R. Chaudhuri[1], Cynthia M. Rodenburg[2], Mario O. Santiago[3] and Beatrice H. Hahn[2,3]

[1]*Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK*
[2]*Departments of Medicine and* [3]*Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA*

In the absence of direct epidemiological evidence, molecular evolutionary studies of primate lentiviruses provide the most definitive information about the origins of human immunodeficiency virus (HIV)-1 and HIV-2. Related lentiviruses have been found infecting numerous species of primates in sub-Saharan Africa. The only species naturally infected with viruses closely related to HIV-2 is the sooty mangabey (*Cercocebus atys*) from western Africa, the region where HIV-2 is known to be endemic. Similarly, the only viruses very closely related to HIV-1 have been isolated from chimpanzees (*Pan troglodytes*), and in particular those from western equatorial Africa, again coinciding with the region that appears to be the hearth of the HIV-1 pandemic. HIV-1 and HIV-2 have each arisen several times: in the case of HIV-1, the three groups (M, N and O) are the result of independent cross-species transmission events. Consistent with the phylogenetic position of a 'fossil' virus from 1959, molecular clock analyses using realistic models of HIV-1 sequence evolution place the last common ancestor of the M group prior to 1940, and several lines of evidence indicate that the jump from chimpanzees to humans occurred before then. Both the inferred geographical origin of HIV-1 and the timing of the cross-species transmission are inconsistent with the suggestion that oral polio vaccines, putatively contaminated with viruses from chimpanzees in eastern equatorial Africa in the late 1950s, could be responsible for the origin of acquired immune deficiency syndrome.

Keywords: human immunodeficiency virus; simian immunodeficiency virus; evolution; cross-species transmission; oral polio vaccine; non-synonymous-to-synonymous rate ratio

1. INTRODUCTION

The viruses that cause acquired immune deficiency syndrome (AIDS), human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2), are lentiviruses. Their closest relatives, found infecting other primates, are termed simian immunodeficiency viruses (SIVs) although as far as is known these viruses do not normally cause disease in their natural hosts. More than 20 species of primates, all from sub-Saharan Africa, are known to harbour SIV (Hahn *et al*. 2000) and at least some of these species exhibit very high rates of infection, with diverse but species-specific viruses. In addition, a number of closely related species of African monkeys are infected by closely related SIVs (Beer *et al*. 1999), suggesting that at least some African primates have been infected with SIV for a very long time (Hahn *et al*. 2000). In contrast, human infection by HIV-1 and HIV-2 seems a relatively recent phenomenon. For both HIV-1 and HIV-2, the greatest diversity of viral strains, as well as the highest rates of infection, occurs in sub-Saharan Africa (Peeters & Sharp 2000). These data provide compelling evidence that the HIVs have arisen through cross-species transmission from primates in Africa. Phylogenetic analyses indicate that HIV-1 and HIV-2 fall into two quite different SIV lineages (figure 1) and so have distinct origins.

Strains of SIV closely related to HIV-2 have been isolated from sooty mangabeys (*Cercocebus atys*) and several macaques (three different *Macaca* species). Only a few macaques, all in captivity in North America, have been found to carry these viruses; these species are not naturally infected with SIV in the wild in Asia. In contrast, SIVsm has been isolated from wild sooty mangabeys in western Africa (Chen *et al*. 1996). HIV-2 is only endemic in western Africa, and it seems clear that SIVsm has been transmitted to humans there, as well as to macaques in captivity. Detailed consideration of the phylogenetic relationships among SIVsm and HIV-2 strains indicates that cross-species transmission has occurred on multiple occasions (Gao *et al*. 1992, 1994; Chen *et al*. 1997).

The only strains of SIV closely related to HIV-1 have been isolated from chimpanzees (*Pan troglodytes*). The first of these SIVcpz strains was discovered more than ten years ago (Peeters *et al*. 1989; Huet *et al*. 1990), but until recently considerable doubt remained as to whether chimpanzees were the natural reservoir of these viruses, because only very few infected animals had been found. There remained the possibility that both humans and chimpanzees became infected from some third species, an as yet unidentified African monkey. However, with the at

*Author for correspondence (paul@evol.nott.ac.uk).

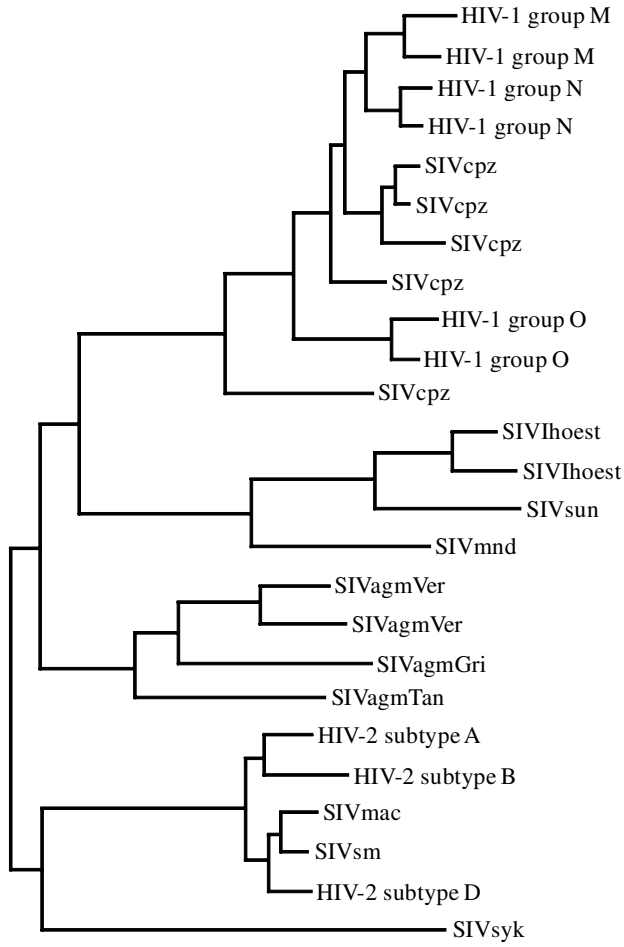

Figure 1. Phylogenetic relationships among representative primate lentiviruses, derived from maximum-likelihood analysis of Pol protein sequences. For HIV-1 the group of origin, and for HIV-2 the subtype, are indicated. SIV strains have a suffix denoting their species of origin: cpz, chimpanzee; lhoest, l'Hoest's monkey; sun, sun-tailed monkey; mnd, mandrill; agm, African green monkey; Ver, vervet; Gri, grivet; Tan, tantalus; mac, macaque; sm, sooty mangabey; syk, Sykes' monkey. See Hahn *et al.* (2000) for more details.

least partial characterization of SIVs from numerous additional primate species (Hahn *et al.* 2000), the list of potential candidates for the elusive third species has shrunk considerably. Furthermore, within the last two years, the characterization of additional strains of SIVcpz (Gao *et al.* 1999; Corbet *et al.* 2000), as well as the discovery of a new strain of HIV-1 even more closely related to SIVcpz (Simon *et al.* 1998), has strongly tipped the balance in favour of chimpanzees being the immediate source of HIV-1.

But how, where and when did the cross-species transmissions that gave rise to the human AIDS viruses, and in particular HIV-1, occur? With respect to where and when, molecular evolutionary analyses provide the clearest evidence to answer these questions. With respect to the question how, molecular evolutionary analyses cannot provide an answer, but do at least seem to rule out one proposed route.

2. HIV-1 DIVERSITY

First it is necessary to review briefly the diversity of HIV-1. Due to its high replication rate and the error-prone nature of its reverse transcriptase, HIV-1 evolves at an extraordinarily fast rate (Li *et al.* 1988; Korber *et al.* 1998), and has accumulated considerable diversity since its arrival in humans. Phylogenetic analyses indicate that the various strains of HIV-1 fall into three groups (Peeters & Sharp 2000). All of the first characterized strains, as well as most of those subsequently described, fall into the 'main' group, group M, responsible for the vast majority of AIDS cases worldwide. This group has been further subdivided into numerous approximately equidistant clades, termed subtypes A–K. Highly divergent strains, forming the 'outlier' group, group O, were first described around ten years ago (De Leys *et al.* 1990). These strains are most commonly found in western equatorial Africa. Most recently, a small number of viruses have been described from Cameroon that belong to a third group (termed N, or 'non-M, non-O') (Simon *et al.* 1998). Since these three groups are interspersed among lineages of SIVcpz, they must each have arisen from a separate cross-species transmission (Gao *et al.* 1999; Hahn *et al.* 2000).

A second important source of genetic diversity in primate lentiviruses is recombination (Sharp *et al.* 1995). The effect of this is most pronounced when a single individual is infected by diverse strains from more than one source. In regions where different subtypes co-circulate, 10% or more of HIV-1 group M strains are inter-subtype recombinants (Robertson *et al.* 1995; McCutchan 1998; Peeters & Sharp 2000) and recently recombinants of groups M and O have been identified (Takehisa *et al.* 1999; Peeters *et al.* 1999). Interestingly, group N is also descended from a recombinant strain (Gao *et al.* 1999; Corbet *et al.* 2000); in this case the recombination event must have occurred in chimpanzees, some time before the transmission of this lineage to humans.

3. ORIGINS OF HIV-1: HOW?

Chimpanzees (as well as sooty mangabeys, and many other species of primate) are often hunted and butchered for food and sold at major city markets as so called 'bush-meat' (Robinson *et al.* 1999). Direct exposure of humans to animal blood or mucosal secretions as a result of hunting and field dressing, or other activities such as consumption of uncooked contaminated meat, would appear to provide the simplest and most plausible explanation for the cross-species transmission of SIVs to become HIVs. However, iatrogenic (i.e. human-induced) modes of cross-species transmission have also been proposed. Among these, the idea that AIDS viruses were transferred from their natural hosts to humans via contaminated preparations of oral polio vaccine (OPV) has recently gained the most widespread attention.

The potential OPV–AIDS link has been mooted on a number of occasions in the past (see, for example, Curtis 1992; Elswood & Stricker 1993) but has only recently been expounded at great length in a book entitled *The river* by the journalist Edward Hooper (Hooper 1999). It is proposed that the various forms of both HIV-1 and

HIV-2 probably had their origins in polio vaccine preparations from different laboratories. In particular, the cause of the HIV-1 group M pandemic has been suggested to be the polio vaccine developed by a team led by Dr Hilary Koprowski and administered to roughly one million people in the Belgian Congo, Rwanda and Burundi in the late 1950s. This research group had access to a primate facility at Camp Lindi near Stanleyville (now Kisangani), where locally caught chimpanzees were used for polio vaccine safety testing as well as other research projects. Hooper suggests that kidneys harvested from these apes were used for polio vaccine production in the USA and/or elsewhere, and that kidneys derived from SIVcpz-infected chimpanzees contaminated these OPV preparations (for more detail see Hooper, this issue). While there is no direct evidence to support this scenario, the circumstantial evidence has been presented sufficiently persuasively to gain some attention, particularly in the lay press (for example, see Ridley 2000).

Various independent lines of evidence, ranging from serious flaws in the proposed epidemiological linkage of OPV and early AIDS cases, to an almost certain loss of SIV/HIV viability due to the process of vaccine production, undermine the OPV–AIDS scenario. Moreover, testimony from the investigators involved, as well as independent experimental data from the analysis of remaining vaccine samples (which at the very least represent a random spot check), argues strongly that chimpanzee kidneys were not used in the vaccine preparation process, and so could not have contaminated it (see other papers in this issue). Our focus is the molecular evolutionary evidence concerning where and when the M group of HIV-1 arose, which is clearly also not consistent with the OPV route. A number of the conclusions from molecular evolutionary analyses have been misinterpreted (Hooper 1999; 2000*a*,*b*) and we present the evidence here in greater detail.

4. ORIGINS OF HIV-1: WHERE?

The source of HIV-1, the common chimpanzee *P. troglodytes*, is distributed across western and equatorial Africa. Recent genetic evidence, largely from analyses of mitochondrial DNA (mtDNA) sequences, indicates that chimpanzees can be divided into at least two, and perhaps four, distinct subspecies (Morin *et al*. 1994; Gonder *et al*. 1997; Gagneux *et al*. 1999). Thus chimpanzees from western Africa, termed *P. t. verus*, are quite distinct from those in equatorial Africa, termed *P. t. troglodytes* and *P. t. schweinfurthii*. As yet no example of SIVcpz infection of *P. t. verus* has been recorded. This may explain the apparent absence of SIVcpz infection among captive chimpanzees outside Africa, since those animals were largely exported from western Africa. There is also some evidence that chimpanzees from Nigeria, which have been termed *P. t. vellerosus*, are genetically distinct. The only reported instance of SIVcpz infection of *P. t. vellerosus* was in an animal held in the same facility as an infected *P. t. troglodytes*: viruses from the two chimpanzees were very closely related, indicating that one animal (presumably the *P. t. vellerosus*) was the recipient of a cage transmission (Corbet *et al*. 2000).

It has long been thought that chimpanzees from equatorial Africa can be divided into two subspecies, the Central chimpanzee *P. t. troglodytes* from western equatorial Africa (Cameroon, Equatorial Guinea, Gabon and Congo-Brazzaville) and the Eastern chimpanzee *P. t. schweinfurthii* from Central Africa (northern and eastern Congo, Rwanda, Burundi and western Uganda and Tanzania). From mtDNA analyses, this separation is not so clear, in that chimpanzees from the two putative subspecies do not exhibit reciprocal monophyly: Eastern chimpanzees cluster within the Central chimpanzee radiation. Nevertheless, the extent of mtDNA differentiation is such that it seems that most individuals of *P. t. schweinfurthii* can be identified with some certainty. Four of the seven known SIVcpz strains clearly came from Central chimpanzees: two from Gabon (Peeters *et al*. 1989), and two from Cameroon (Corbet *et al*. 2000). A fifth came from a wild-born animal held in a primate centre in the USA that was identified as a Central chimpanzee on the basis of mtDNA analyses (Gao *et al*. 1999). Excluding the *P. t. vellerosus* example, the only SIVcpz strain (SIVcpzAnt) not isolated from a Central chimpanzee came from a wild-caught animal of imprecise Congolese origin, intercepted by Belgian customs officers after illegal export from Kinshasa (Peeters *et al*. 1992); this chimpanzee was identified as belonging to *P. t. schweinfurthii*, again on the basis of mtDNA analyses (Gao *et al*. 1999). Since all three groups of HIV-1 are more closely related to the SIVcpz strains from Central chimpanzees, we concluded that the cross-species transmissions giving rise to HIV-1 all occurred in western equatorial Africa (Gao *et al*. 1999). This is not consistent with the OPV–AIDS idea, since the chimpanzees at Camp Lindi were almost entirely captured in the north-eastern Congo (Courtois 1966), in the middle of the Eastern chimpanzee range.

In response to the conclusion that Central chimpanzees in western equatorial Africa were the source of all three groups of HIV-1, Hooper has written: 'Whereas it [the SIVcpz from Central chimpanzees] is an entirely convincing parent for HIV-1 group N ... and a slightly less convincing parent for the other minor variant, group O ... it represents a far from convincing parent for the pandemic variant, group M' (Hooper 2000*a*, p. 867). However, consideration of the phylogenetic relationships among the SIVcpz strains and the three HIV-1 groups shows that this statement is not supported by existing data. Both groups M and N lie within the radiation of SIVcpz strains from Central chimpanzees (figure 2). As discussed above, group N is inferred to be descended from a recombination event among divergent lineages, because its position within phylogenetic trees varies according to the region of genome analysed. Presumably it is the position of group N within the phylogeny derived from Env protein sequences (figure 2*b*) that has convinced Hooper that this group was derived from a Central chimpanzee SIVcpz. In phylogenies derived from the 5′ half of the genome (such as the *Pol* gene; figure 2*a*) group N clusters more closely with group M than with any SIVcpz strain, although both groups M and N lie within the radiation of SIVcpz strains from Central chimpanzees. In this 5′ region, groups M and N are equidistantly related to SIVcpz from Central chimpanzees, and so if it is

acknowledged that these are 'an entirely convincing parent for HIV-1 group N', then they must represent an equally convincing source for group M also. It is also worth noting that while group M strains have spread across all of Africa, now having their highest frequency in the south of the continent and thus outside the range of chimpanzees, the greatest diversity of group M strains is found in western equatorial Africa, i.e. in Kinshasa rather than further east in the Democratic Republic of Congo (Vidal *et al*. 2000), consistent with this being their region of origin. (Kinshasa is outside the range of chimpanzees, but is close to the area inhabited by Central chimpanzees, and is by far the largest city in the region.)

Among the three groups of HIV-1, the one that does not fall within the radiation of SIVcpz from Central chimpanzees is group O. Thus on phylogenetic grounds alone, the origin of group O is less certain, although the strains of SIVcpz from Central chimpanzees are the most closely related known viruses for this group also. Since group O strains are most common in, and largely restricted to, Cameroon and neighbouring countries, it is most likely that they too arose from a cross-species transmission in this region.

All currently known SIVcpz strains were derived from chimpanzees that were captured at a young age, most likely (and this is known in several cases) after their parents were killed. Since collection of blood from wild ape populations is not feasible, the prevalence, geographical distribution and extent of diversity of SIVcpz in the wild remain unknown. To investigate this, we have recently developed non-invasive methods to detect SIVcpz antibodies and viral nucleic acids in chimpanzee urine and faecal samples collected in the wild (Santiago *et al*. 2001). The sensitivity of antibody detection was tested in captive chimpanzees of known SIVcpz or HIV-1 infection status and found to be 100% for urine and greater than 60% for faeces; in each instance the specificity was 100%. The sensitivity of polymerase chain reaction amplification of viral RNA from faeces of SIVcpz-infected chimpanzees was greater than 50%. Using these assays, we have begun to test chimpanzees from wild communities in western and eastern Africa. While all of the 28 West African chimpanzees tested were negative for SIVcpz antibodies and nucleic acids, 1 out of 30 East African chimpanzees was strongly positive for urine SIVcpz antibodies. This infected individual was one of six chimpanzees sampled from a larger community. Although viral sequences are not yet available, the profile of the Western blot analysis strongly suggests that this strain is not similar to either HIV-1 or SIVcpz from Central chimpanzees. These data demonstrate the feasibility of screening wild chimpanzee populations using entirely non-invasive means and represent, we believe, the first detection of SIVcpz infection in eastern Africa. Prevalence and molecular epidemiological data of naturally occurring SIVcpz infections (including in the vicinity of Kisangani) can be expected to be forthcoming soon.

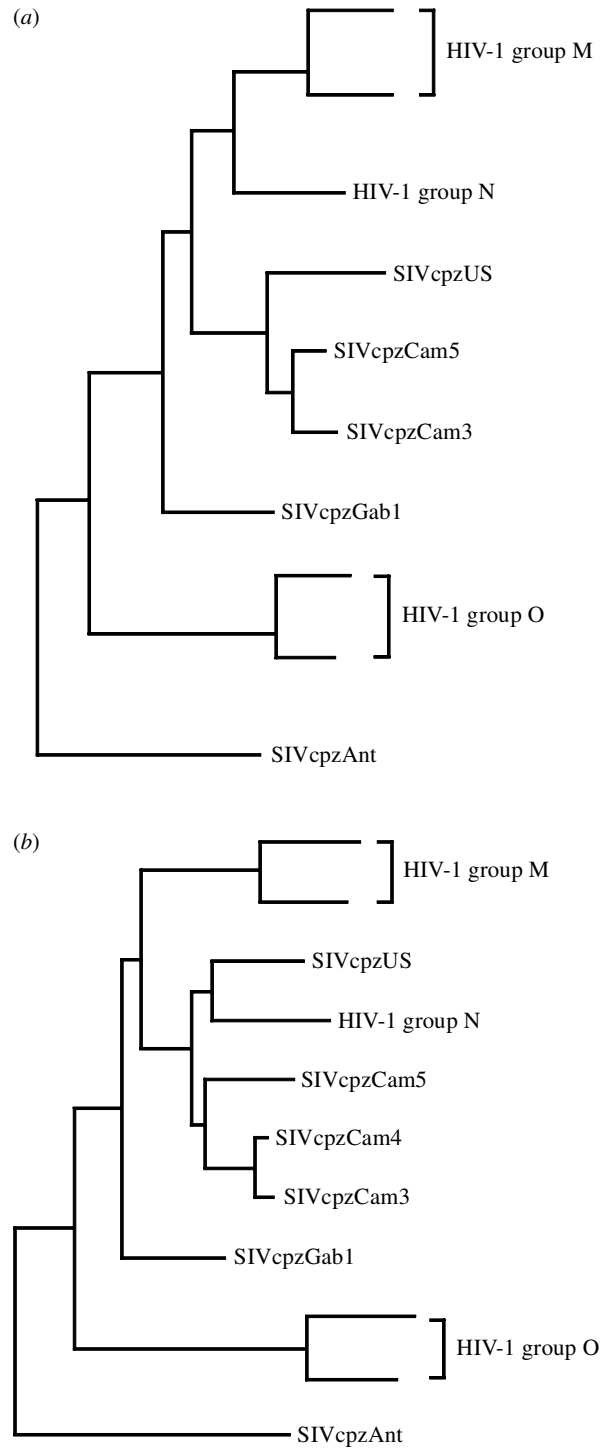

Figure 2. Phylogenetic relationships among SIVcpz strains and representatives of the three groups (M, N and O) of HIV-1, derived from maximum-likelihood analysis of (*a*) Pol and (*b*) Env protein sequences. All SIVcpz strains came from animals identified as *P. t. troglodytes* (Central chimpanzees), except Cam4 (from *P. t. vellerosus*) and Ant (from *P. t. schweinfurthii*). The position of HIV-1 group N differs in the two trees, indicating that it is a descendant of a recombination event.

5. ORIGINS OF HIV-1: WHEN?

Consideration of the phylogenetic relationships among HIV-1 and SIVcpz strains (figure 2) clearly indicates that each group of HIV-1 represents a separate cross-species transmission event, and that it is most parsimonious to invoke only one such event per group (figure 3*a*). Thus, if the date of the common ancestor within each group can be estimated, this can place a limit (the most recent

bound) on the time-scale of when the cross-species transmissions occurred. Attempts to date these events have focused on the M group of HIV-1. Early attempts to place a time-scale on the evolution of the primate lentiviruses gave wildly divergent results (see Sharp & Li 1988). However, after estimating the rate of the molecular clock of HIV-1 it was suggested that the common ancestor of (what is now known as) the M group existed around 1960 (Li *et al*. 1988). This date seemed to fit with knowledge of the early AIDS epidemic, and with the earliest known (retrospectively identified) antibody-positive serum sample obtained from an individual in Léopoldville (now Kinshasa) in 1959 (Nahmias *et al*. 1986). This impression of the time-scale became established (for example, Myers & Korber 1994), and remained largely unchallenged for nearly ten years. This time-scale would have been consistent with the OPV–AIDS hypothesis.

However, it has become apparent in recent years that the time-depth of the M group radiation has been severely underestimated. Two lines of evidence point to this. First, sequences were obtained from the 1959 sample. When this virus (ZR59) was placed within the HIV-1 evolutionary tree, it was found to lie within the M group on the subtype D lineage, after its split from subtype B (Zhu *et al*. 1998), with the clear implication that the common ancestor of the entire M group must have existed rather earlier. Second, it became apparent that the estimate of 1960 had been derived from an analysis that assumed an oversimplistic model of nucleotide sequence evolution. In recent years more complex models have been developed that take into account heterogeneous rates of evolution at different sites within genes. This is important because HIV-1 sequences can be seen to evolve in this way (Leitner *et al*. 1997). Ignoring this rate heterogeneity leads to an underestimation of the true amount of divergence among sequences and thus an underestimate of the time-depth of the phylogeny (Sharp *et al*. 2000). Incorporating gamma-distributed among-site rate variation into the molecular clock analysis pushed the estimate of the common ancestor of the M group back from 1960 to before 1940 (Sharp *et al*. 2000).

Korber *et al*. (2000*a*) have performed the most complete analysis of this question. Analysing envelope gene sequences of HIV-1 isolates from more than 150 individuals, and using a maximum-likelihood model of sequence evolution allowing individual sites to evolve at different rates, they estimated the common ancestor of the M group at 1931, with a confidence interval of 1915–1941. Corroboration of the validity of the approach used came with the correct prediction of the date of origin of the ZR59 sample. (See also Yusim *et al*., this issue.)

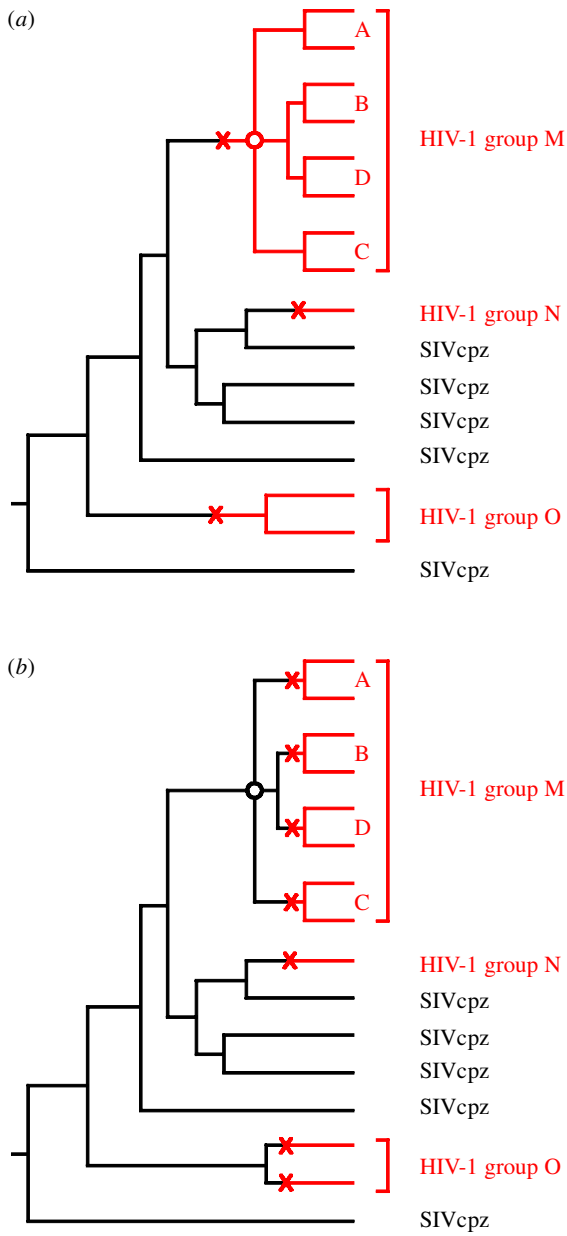

Figure 3. Alternative hypotheses concerning the points of chimpanzee-to-human transmission of SIVcpz to form HIV-1. The common ancestor of HIV-1 group M, which has been dated to pre-1940, is denoted by a circle. Only four subtypes of the M group are illustrated; at least ten have so far been described. These subtypes are approximately equidistantly related, except for subtypes B and D, which are consistently clustered. (*a*) Early origin, whereby each group of HIV-1 was derived from a single cross-species transmission. (*b*) Late origin, where (in particular) the different subtypes of HIV-1 group M were derived from different strains of SIVcpz, as required by the OPV–AIDS hypothesis.

6. DATING AN ABSTRACTION?

If the common ancestor of the HIV-1 M group was a human virus, i.e. occurred after the chimpanzee-to-human transmission, then the dating results described above clearly contradict the idea that transmission was due to contaminated OPV, since that would have occurred in the late 1950s. In response, the leading proponent of the OPV–AIDS idea has argued that 'Dr Korber's work ... dates an abstraction. It does not tell us if the 1931 "last common ancestor" of HIV-1(M) was a chimp virus or a human virus' (Hooper 2000*b*). In other words, to sustain the OPV–AIDS hypothesis, Hooper has now suggested that the last common ancestor of the M group was still residing in chimpanzees, and therefore that multiple strains of SIVcpz were transmitted to humans (figure 3*b*). Various lines of evidence from the available data indicate that this suggestion is quite implausible.

If it is argued that the original diversification of the M group occurred in chimpanzees, the distinct lineages

within HIV-1 group M would have arisen in one of two ways. They could represent either diverse members of a quasi-species from a single infected chimpanzee or viruses from numerous infected chimpanzees. Note that an intermediate scenario, where the multiple lineages came from two (or a few) quasi-species, would not be consistent with the data because there are not two (or a few) distinct clades within the M group.

First, how many distinct viruses would need to have been transmitted? To be consistent with the time-scale, there must have been at least one for each subtype, of which there are currently ten or more (Peeters & Sharp 2000). In addition, there are other distinct lineages, as divergent as the current subtypes (Vidal *et al*. 2000), which have not been given subtype status because only one strain has thus far been characterized (Robertson *et al*. 2000). Furthermore, for whatever number of distinct viruses are hypothesized to have been successfully transmitted, it must be expected that there would have to have been numerous additional strains that, simply by chance, did not become established in the human population or at least have no detected descendants. Thus, one would have to postulate some very large number of divergent viruses (perhaps as many as 50?) in the putatively contaminated vaccines.

Under the single donor scenario, this large number of members of a quasi-species, all showing approximately 25 years' worth of divergence from one another, would need to have been transmitted. This is implausible for two reasons. First, the vast majority of the chimpanzees at Camp Lindi were under four years old (Korber *et al*. 2000*b*). Second, even with a chimpanzee that had been infected for 25 years, a sample from the viral quasi-species would contain strains that had diverged at various times over this period, and perhaps none that had diverged as long as 25 years ago; thus to expect that one would sample only strains that had diverged from one another 25 years earlier is simply unrealistic. Alternatively, under the multiple donor scenario, we would need to invoke a population of chimpanzees with SIVs all dating back to a common ancestor only (but no less than) 25 years earlier. Again, the phylogenetic data for available SIVcpz strains argue strongly against this. Thus this time-gap of around 25 years between the estimated date of the common ancestor of the M group and the period when OPV was administered is much too long for the single donor scenario, but much too short for the multiple donor scenario.

7. RECOMBINATION

Retroviruses are highly recombinogenic. There is widespread evidence of recombination among HIV-1 strains, both within quasi-species (Vartanian *et al*. 1991; Howell *et al*. 1991; Groenink *et al*. 1992) and involving discrete strains after multiple infections (Robertson *et al*. 1995, 1997; McCutchan 1998). It must be expected that this also occurs among strains of SIVcpz. If, as now hypothesized by Hooper (2000*b*), the M group of HIV-1 arose through the transmission of multiple strains of SIVcpz, derived from either a single or multiple chimpanzees, we must expect that those strains would have been undergoing recombination prior to their arrival in humans. Thus, if the putative OPV contamination involved diverse members of a single quasi-species, those strains would have been undergoing recombination in the infected chimpanzee. Alternatively, if the putative OPV contamination involved diverse strains from multiple chimpanzees it would imply a sufficiently high rate of infection that, combined with chimpanzee behaviour patterns (Goodall 1986; Wrangham *et al*. 1992), should have led to instances of multiple infection and as a consequence recombination among divergent SIVcpz strains. Indeed this is known to have occurred among strains of SIVcpz from western equatorial Africa, since HIV-1 group N viruses are descended from a virus that was a recombinant of divergent SIVcpz strains.

However, there is little or no evidence of recombination having occurred during the early divergence of the lineages giving rise to the subtypes within HIV-1 group M. Recombination events deep within the M group phylogeny would be hard to identify (Schierup & Hein 2000), but the very fact that discrete subtypes can be identified indicates that recombination must have been rare during this period of early diversification within the M group. The recombination events that have been identified have all involved strains subsequent to the establishment of the subtypes. This is as expected if the common ancestor of the M group existed after the transmission of SIVcpz to humans. During the early period of diversification of the M group, the number of infected people would have been quite small and multiple infection would have been unlikely, so that there would have been little opportunity for recombination among divergent strains. In contrast, the apparent scarcity of recombination during the pre-subtype phase of M group diversification would be quite unexpected if these viral lineages were still in chimpanzees.

It has been remarked that the three earliest examples of HIV-1 group M for which sequence data are available, namely ZR59, Z321 and MAL, were all multiply recombinant and that the recombination events may have occurred before the transmission to humans (Hooper 2000*a*, p. 1062). However, there is no evidence for this. First, two of these isolates are not particularly 'early' in the context of the time-scale of the subtypes: Z321 was isolated in 1976 from a Zairian woman who had lived in Kinshasa between 1972 and 1976 where she was considered a 'free woman' (Nzilambi *et al*. 1988), while MAL was probably contracted from a blood transfusion in Zaire in 1981 (Alizon *et al*. 1986). Second, and more important, the analyis of the only really early sample, ZR59 from 1959, provided no evidence for the suggestion that it was a recombinant (Zhu *et al*. 1998).

Finally, widespread recombination among a group of sequences can affect the estimated time of their common ancestor. In the context of dating the common ancestor of HIV-1 group M, it has been suggested that it is unclear 'whether recombination would make the introduction date earlier or later' (Hooper 2000*a*, p. 870). This statement is strange: it is clear that recombination would make the date of the common ancestor seem more recent (Schierup & Hein 2000). Korber *et al*. (2000*a*) removed identifiable inter-subtype recombinant sequences from their analyses, but if there were any such strains remaining in the data set the consequence would be that

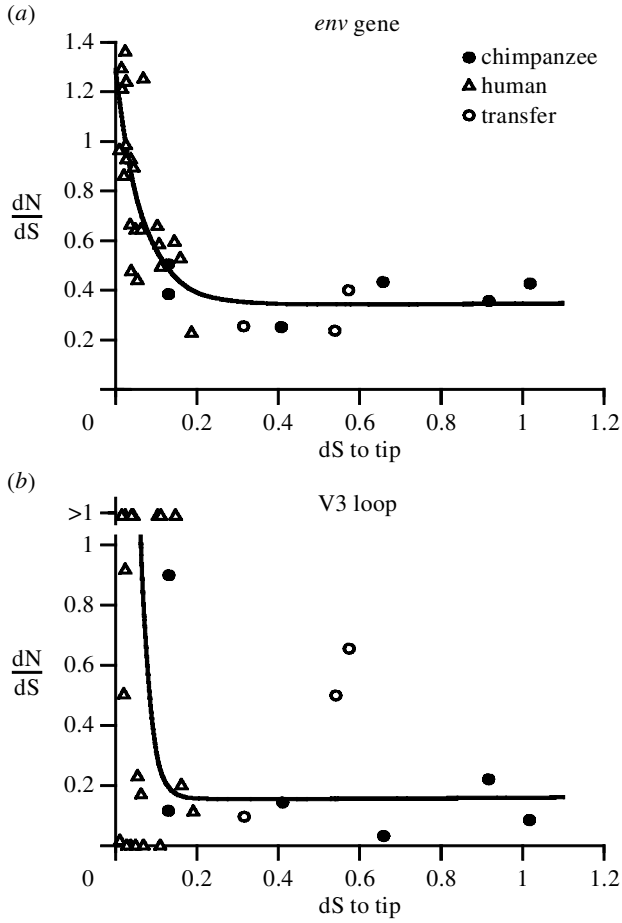

Figure 4. dN/dS on different branches within the SIVcpz–HIV-1 phylogeny, as a function of the depth of the branches within the tree (dS to tip). Values of dN and dS were estimated using PAML (Yang 1997); dS to tip was calculated as the average distance from the midpoint of the branch to the tips of descendant branches. Black circles denote branches where the viruses were evolving in chimpanzees, open triangles denote branches where the virus was evolving in humans, and open circles denote the branches where cross-species transmission occurred, under the hypothesis of single transmissions giving rise to each of the three groups of HIV-1 (figure 3*a*). (*a*) Derived from the entire *env* gene. (*b*) Derived from the V3 loop encoding region of the *env* gene. Note that the dN/dS axes are drawn to different scales in (*a*) and (*b*).

the true date of the common ancestor of the M group would be even earlier than the estimate of 1931.

8. RATES OF NON-SYNONYMOUS SUBSTITUTION

The pattern of nucleotide substitution on the early branches within the M group of HIV-1 could also be informative about when, i.e. where on the evolutionary tree, transmissions from chimpanzees to humans occurred. The degeneracy of the genetic code implies that, within protein-coding sequences, nucleotide substitutions can be categorized as non-synonymous (or synonymous) depending on whether they affect (or do not affect) the amino-acid sequence encoded. It has often been suggested that the rate of non-synonymous substitutions might increase after cross-species transmission of a pathogen, reflecting adaptation to the new host. In contrast, synonymous substitutions are expected to be largely neutral, and reflect only the underlying rate of mutation and replication. Therefore the relative rate of non-synonymous substitution, dN (also referred to as K_A), can be calibrated by comparison with the rate of synonymous substitution, dS (or K_S), using the dN/dS (or K_A/K_S) ratio.

Using such an approach, Shpaer & Mullins (1993) reported a correlation between rates of protein evolution and pathogenicity. Specifically, they found higher dN/dS ratios in HIV-1 and SIVmac than in SIVsm and SIVagm, for the *gag* gene and the segment of *env* encoding the surface glycoprotein (gp120). Like HIV-1, SIVmac is due to a recent cross-species transmission (although the source of SIVmac was SIVsm from sooty mangabeys, as for HIV-2), and SIVmac causes a severe AIDS-like illness in macaques. In contrast, SIVagm resembles SIVsm insofar as it represents a natural infection of African primates (in this case, African green monkeys) and is not known to cause any disease. However, dN in HIV-2 were not found to be substantially higher than in SIVsm. Rates of evolution in SIVcpz were not investigated at that time. Mindell (1996) reported a lower dS/dN ratio (i.e. a higher dN/dS) for *gag* and *env* (gp120) among HIV-1 group M strains than in SIVcpzGab1. These reports suggest that we can look at dN/dS ratios on individual branches within the SIVcpz/HIV-1 phylogeny, to see where within the evolution of HIV-1 the ratio has increased, and so infer which branches involved chimpanzee-to-human transmission.

We have investigated this recently, using a maximum-likelihood approach (Yang 1998) to estimate the values of the dN/dS ratio on each branch within the phylogenetic tree. For each of the *gag*, *pol* and *env* genes, we have found that dN/dS values vary among branches, but that this is not simply correlated with whether the branch reflects evolution of the virus in chimpanzees or in humans. That is, the dN/dS values are not all higher in HIV-1 than in SIVcpz. Rather, the higher dN/dS values are all found on branches near the tips of the tree, i.e. recent branches can have high dN/dS. The results for the *env* gene are shown (figure 4*a*): for the *gag* and *pol* genes different values of dN/dS were seen, presumably reflecting different levels of constraint on the products of these genes, but the same overall trend of increased dN/dS only on recent branches was observed. These results seem to indicate the short-term accumulation of slightly deleterious, non-synonymous substitutions that become eliminated over time (perhaps, for example, at the point of transmission between individuals). These results are interesting, but imply that this approach does not immediately reveal when the viruses were subject to cross-species transmission.

In our earlier attempts to identify viral genetic changes associated with transmission from chimpanzees to humans, the principal region identified was the part of the envelope gene encoding the V3 loop (Van den Haesevelde *et al*. 1996; Hahn *et al*. 2000). The V3 loop lies on the outside of gp120 and forms part of the binding site used by the virus to engage a cell surface co-receptor during the fusion process (Choe *et al*. 1996; Wyatt & Sodroski 1998). V3 also represents a neutralizing antibody binding site (Parren *et al*. 1999). Hence the V3 loop is under strong *in vivo* selection pressures related to alterations in viral tropism (co-receptor

choice) and immune evasion (neutralizing antibody escape). It is thus not surprising that amino-acid sequence diversity in this region is much higher among HIV-1 strains than among SIVcpz strains (including SIVcpzAnt), even though in trees based on whole protein sequences the diversity among SIVcpz strains is the higher (figure 2). When the dN/dS analysis reported above was applied to V3 loop sequences only, the same overall trend was seen as for entire *gag*, *pol* and *env* gene sequences (figure 4*b*). However, there are two branches showing extremely high dN/dS values (around 0.6) given their depth within the tree (dS to tip values greater than 0.5). These unusually high values come from the branches prior to the common ancestor of group M and the common ancestor of group O. These results suggest that the accelerated rate of V3 evolution, and thus the altered evolutionary pressures associated with a change of host, occurred prior to the common ancestors of group M and O; i.e. that the chimpanzee-to-human transmissions occurred prior to the common ancestors of groups M and O.

9. CONCLUSIONS

HIV-1 groups M, N and O represent three distinct cross-species transmissions of SIVcpz, with all of the evidence pointing to western equatorial Africa as the location where these events took place. The common ancestor of HIV-1 group M existed before 1940, with all of the evidence indicating that this virus was already infecting humans at that point. Both of these observations contradict the suggestion that the human viruses arose via contaminated OPVs.

If 'natural' transfer can occur through human exposure to infected primate blood or excretions during hunting or other activities, the obvious question that remains is why the AIDS epidemic arose only in the latter half of the 20th century. In this context it is important to distinguish between two stages in the advent of the epidemic. The first stage is the transmission of SIV to a human. Given the evidence for several transfers of both SIVcpz (giving rise to the three groups of HIV-1) and SIVsm (leading to the multiple subtypes of HIV-2), it seems likely that cross-species transmission has occurred on a multitude of occasions in the past. The second stage is the successful establishment of an epidemic by the newly arrived virus. It is this second stage that seems new, and most likely reflects changes in population structure and behaviour in Africa during the 20th century and perhaps medical interventions that provided the opportunity for rapid human-to-human spread of the virus (Chitnis *et al*. 2000). It is clear that even in recent times, different transfers of SIVcpz and SIVsm (the different groups of HIV-1 and the subtypes of HIV-2) have succeeded to very different extents, ranging from subtypes of HIV-2 that have only ever been described in a single individual to the M group of HIV-1 that has already infected more than 50 million people worldwide. Ruling out OPV as a viable source of human infection is important because it means that we still do not understand which viral, host and/or environmental factors played a role in sparking the AIDS pandemic.

This work was supported by grants NO1 AI 85338, RO1 AI 44596 and P20 AI27767 from the US National Institutes of Health to B.H.H., and by a studentship from the UK Biotechnology and Biological Sciences Research Council to R.R.C.

REFERENCES

Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. 1986 Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**, 63–74.

Beer, B. E., Bailes, E., Sharp, P. M. & Hirsch, V. M. 1999 Diversity and evolution of primate lentiviruses. In *Human retroviruses and AIDS 1999: a compilation and analysis of nucleic acid and amino acid sequences* (ed. C. L. Kuiken, B. Foley, B. Hahn, B. Korber, F. McCutchan, P. A. Marx, J. W. Mellors, J. I. Mullins, J. Sodroski & S. Wolinsky), pp. 460–474. Los Alamos, NM: Los Alamos National Laboratory.

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.

Chen, Z. (and 11 others) 1997 Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J. Virol.* **71**, 3953–3960.

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

Choe, H. (and 12 others) 1996 The beta chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135–1148.

Corbet, S. (and 11 others) 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.

Courtois, G. 1966 Sur la réalisation d'une singerie de chimpanzes au Congo. In *Symposium international sur l'avenir des animaux de laboratoire, Lyon, 18–20 September 1966*, pp. 235–244. Lyon, France: Institut Mérieux.

Curtis, T. 1992 The origin of AIDS. *Rolling Stone* 19 March.

De Leys, R. (and 15 others) 1990 Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J. Virol.* **64**, 1207–1216.

Elswood, B. F. & Stricker, R. B. 1993 Polio vaccines and the origin of AIDS. *Res. Virol.* **144**, 175–177.

Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P. A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O. A. & Woodruff, D. S. 1999 Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl Acad. Sci. USA* **96**, 5077–5082.

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

Gao, F. (and 11 others) 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. (and 11 others) 1999 Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436–441.

Gonder, M. K., Oates, J. F., Disotell, T. R., Forstner, M. R. J., Morales, J. C. & Melnick, D. J. 1997 A new west African chimpanzee subspecies? *Nature* **388**, 337.

Goodall, J. 1986 *The chimpanzees of Gombe: patterns of behaviour*. Cambridge, UK: Belknap Press.

Groenink, M., Andeweg, A. C., Fouchier, R. A., Broersen, S., Van der Jagt, R. C., Schuitemaker, H., de Goede, R. E., Bosch, M. L., Huisman, H. G. & Tersmette, M. 1992 Phenotype-associated *env* gene variation among eight related human immunodeficiency virus type 1 clones: evidence for *in vivo* recombination and determinants of cytotropism outside the V3 domain. *J. Virol.* **66**, 6175–6180.

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.

Hooper, E. 1999 *The river: a journey to the source of HIV and AIDS*. London: Allen Lane.

Hooper, E. 2000*a* *The river: a journey to the source of HIV and AIDS*. London: Penguin.

Hooper, E. 2000*b* Search for the origin of HIV and AIDS. *Science* **289**, 1140.

Howell, R. M., Fitzgibbon, J. E., Noe, M., Ren, Z. J., Gocke, D. J., Schwartzer, T. A. & Dubin, D. T. 1991 *In vivo* sequence variation of the human immunodeficiency virus type 1 *env* gene—evidence for recombination among variants found in a single individual. *AIDS Res. Hum. Retroviruses* **7**, 869–876.

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G. & Wain-Hobson, S. 1990 Genetic organisation of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356–359.

Korber, B., Theiler, J. & Wolinsky, S. 1998 Limitations to a molecular clock applied to considerations of the origin of HIV-1. *Science* **280**, 1868–1871.

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

Korber, B., Bhattacharya, T., Theiler, J., Gupta, R., Lapedes, A., Hahn, B., Gao, F., Muldoon, M. & Wolinsky, S. 2000*b* Search for the origin of HIV and AIDS (response). *Science* **289**, 1140–1141.

Leitner, T. A., Kumar, S. & Albert, J. 1997 Tempo and mode of nucleotide substitutions in *gag* and *env* gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *J. Virol.* **71**, 4761–4770.

Li, W.-H., Tanimura, M. & Sharp, P. M. 1988 Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**, 313–330.

McCutchan, F. E. 1998 Global diversity in HIV. In *The evolution of HIV* (ed. K. A. Crandall), pp. 41–101. Baltimore, MD: Johns Hopkins University Press.

Mindell, D. P. 1996 Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees. *Proc. Natl Acad. Sci. USA* **93**, 3284–3288.

Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J. & Woodruff, D. S. 1994 Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**, 1193–1201.

Myers, G. & Korber, B. 1994 The future of human immunodeficiency virus. In *The evolutionary biology of viruses* (ed. S. S. Morse), pp. 249–270. New York: Raven Press.

Nahmias, A. J. (and 11 others) 1986 Evidence for human infection with an HTLV III/LAV-like virus in central Africa, 1959. *The Lancet* **i**, 1279–1280.

Nzilambi, N., De Cock, K. M., Forthall, D. N., Francis, H., Ryder, R. W., Malebe, I., Getchell, J., Laga, M., Piot, P. & McCormick, J. B. 1988 The prevalence of infection with human immunodeficiency virus over a 10-year period in rural Zaire. *New Engl. J. Med.* **318**, 276–279.

Parren, P. W. H. I., Moore, J. P., Burton, D. R. & Sattentau, Q. J. 1999 The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. *AIDS* **13**, S137–S162.

Peeters, M. & Sharp, P. M. 2000 Genetic diversity of HIV-1: the moving target. *AIDS* **14**, S129–S140.

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* **6**, 447–451.

Peeters, M., Liegeois, F., Torimoto, N., Bourgeois, A., Mpoudi, E., Vergne, L., Saman, E., Delaporte, E. & Saragosti, S. 1999 Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J. Virol.* **73**, 7368–7375.

Ridley, M. 2000 The origin of AIDS. *Prospect* **53**, 31–35.

Robertson, D. L., Sharp, P. M., McCutchan, F. E. & Hahn, B. H. 1995 Recombination in HIV-1. *Nature* **374**, 124–126.

Robertson, D. L., Gao, F., Hahn, B. H. & Sharp, P. M. 1997 Intersubtype recombinant HIV-1 sequences. In *Human retroviruses and AIDS 1997: a compilation and analysis of nucleic acid and amino acid sequences* (ed. B. Korber, B. H. Hahn, B. Foley, J. W. Mellors, T. Leitner, G. Myers, F. E. McCutchan & C. Kuiken), pp. III.25–III.30. Los Alamos, NM: Los Alamos National Laboratory.

Robertson, D. L. (and 19 others) 2000 HIV-1 nomenclature proposal. *Science* **288**, 55–57.

Robinson, J. G., Redford, K. H. & Bennett, E. L. 1999 Wildlife harvest in logged tropical forests. *Science* **284**, 595–596.

Santiago, M. O. (and 15 others) 2001 AIDS as a zoonosis: characterizing the primate reservoir. Abstract 286, *8th Conference on Retroviruses and Opportunistic Infections*, 4–8 February 2001, Chicago (http://www.retroconference.org).

Schierup, M. H. & Hein, J. 2000 Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879–891.

Sharp, P. M. & Li, W.-H. 1988 Understanding the origins of AIDS viruses. *Nature* **336**, 315.

Sharp, P. M., Robertson, D. L. & Hahn, B. H. 1995 Cross-species transmission and recombination of 'AIDS' viruses. *Phil. Trans. R. Soc. Lond.* B **349**, 41–47.

Sharp, P. M., Bailes, E., Gao, F., Beer, B. E., Hirsch, V. M. & Hahn, B. H. 2000 Origins and evolution of AIDS viruses: estimating the time-scale. *Biochem. Soc. Trans.* **28**, 275–282.

Shpaer, E. G. & Mullins, J. I. 1993 Rates of amino acid change in the envelope protein correlate with pathogenicity of primate lentiviruses. *J. Mol. Evol.* **37**, 57–65.

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

Takehisa, J., Zekeng, L., Ido, E., Yamaguchi-Kabata, Y., Mboudjeka, I., Harada, Y., Miura, T., Kaptue, L. & Hayami, M. 1999 Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J. Virol.* **73**, 6810–6820.

Van den 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.

Vartanian, J.-P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. 1991 Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779–1788.

Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B. & Delaporte, E. 2000 Unprecedented degree of human immunodeficiency virus type I (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.

Wrangham, R. W., Clark, A. P. & Isabirye-Basuta, G. 1992 Female social relationships and social organization of the Kibale Forest chimpanzees. In *Topics in primatology*, vol. 1. *Human origins* (ed. T. Nishida), pp. 81–98. University of Tokyo Press.

Wyatt, R. & Sodroski, J. 1998 The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science* **280**, 1884–1888.

Yang, Z. 1997 PAML, a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 306–314.

Yang, Z. 1998 Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573.

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.