# Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes* [1]

## I. Introduction

The Acquired Immunodeficiency Syndrome (AIDS) was first recognized in 1981 and has since become a major worldwide epidemic. There is now clear evidence that AIDS is caused by the human immunodeficiency virus (HIV) which binds to the CD4 cell surface receptors found on T-lymphocyte cells (these cells play a role in the regulation of the immune response to invading parasites).[2] HIV is part of a family of viruses called lentiviruses (a subfamily of the retroviruses). Lentiviruses other that HIV have been found in a wide range of nonhuman primates, known as Simian Immunodeficiency Virus (SIV) where a subscript will be used to denote their species of origin.

There are two types of HIV: HIV-1 and HIV-2. HIV-1 is responsible for the global pandemic while HIV-2 has, until recently, been restricted to West Africa and appears to be less virulent in its effects. It is now generally accepted that HIV is a descendant of SIV, that is, HIV represent cross-species infections, or say, is zoonois (the process that certain viruses pass from animals to humans). [3, 4] For HIV-2, a virus (SIVsm) that is genomically indistinguishable and closely related phylogenetically was found in substantial numbers of wild-living sooty mangabey monkeys whose natural habitat coincides with the epicentre of the HIV-2 epidemic (close contact between these monkeys and human is common because the monkeys are hunted for food and kept as pets). Thus the primate reservoir of HIV-2 has been clearly identified as the sooty mangabey [4]. In contrast, the origin of HIV-1 is much less certain. HIV-1 is most similar in sequence and genomic organization to viruses found in chimpanzees (SIVcpz) [3], but a wide spectrum of diversity between HIV-1 and SIVcpz, and apparent low prevalence of SIVcpz infection in wild-living animals, and the presence of chimpanzees in geographic regions of Africa where AIDS was not initially recognized have cast doubt on chimpanzees as a natural host and reservoir for HIV-1.

In *Nature*, February 1999, it was announced that a group of researchers from the University of Alabama had studied frozen tissue from a chimpanzee, came from a sub-group known as *Pan troglodytes troglodytes*. It is claimed by these researchers that this show that all HIV-1 strains known to infect man are closely related to just one of the SIVcpz lineages that found in *Pan troglodytes troglodytes*. These results indicate that *Pan troglodytes troglodytes* may be the primary reservoir for HIV-1, and that the virus at some point crossed species from chimpanzees to human. [1]

## II. Methods and results

Viruses related to HIV-1 have been isolated from the common chimpanzee (*Pan troglodytes troglodytes*), but only three SIVcpz strains have been reported, two from animals wild-caught in Gabon (SIVcpzGAB1, SIVcpzGAB2) and one from a chimpanzee exported to Belgium from Zaire (SIVcpzANT). SIVcpzGAB1 and SIVcpzANT have been sequenced completely, but only 280 base pairs of *pol* sequences are available for SIVcpzGAB2.

In search of HIV-1 reservoir, those researchers identified a fourth chimpanzee with natural infection. This animal (Marilyn) was wild-caught in Africa, exported to the United States as an infant, and used as a breeding female in a primate facility until her death at the age of 26 years. The researchers used the Polymerase Chain Reaction (PCR) to amplify HIV- or SIV- related DNA sequences directly from the frozen (-20°C ) spleen and lymph-node tissue obtained at autopsy in order to characterize the infection responsible for Marilyn's HIV-1 seropositivity (displaying a positive diagnosis to the basic antigen-antibody reaction in vitro). Amplification and sequence analysis of subgenomic *gag* (508 base pairs) and *pol* (766 base pairs) fragments revealed the presence of a virus related to, but distinct from, known SIVcpz and HIC-1 strains. Because virus isolation from the autopsy tissues was unsuccessful, they used PCR to amplify and sequence four overlapping subgenomic fragments that together comprised a complete proviral genome, termed SIVcpzUS. Analysis of potential coding regions revealved the presence of a *vpu* gene (found only in HIV-1 and SIVcpz viruses) [3], in addition to structural and regulatory genes common to all primate lentiviruses.

To determine the evolutionary relationship of SIVcpzUS to other HIV and SIV sequences, they performed distance plot and phylogenetic tree analyses using sequences from the HIV sequence database (http://hiv-web.lanl.gov/HTML/compendium.html). These analyses identified SIVcpzUS unambiguously as a new member of the HIV-1/SIVcpz group of viruses. In Fig 1(a), a phylogenetic tree of full-length *pol* sequences showed that the HIVcpzUS clustered well within this group but was not particularly closely related to any one human or chimpanzee virus (trees based on other coding regions yielded virtually identical topologies). Comparison of the phylogenetic position of SIVcpzUS with those of the other SIVcpz strains showed that SIVcpzUS was considerable more closely related to SIVcpzGAB1 than to SIVcpzANT. In Fig 1(b), diversity plots of full-length (concatenated) protein sequences showed that partial Pol sequences of SIVcpzUS was nearly twice as different from SIVcpzANT as from SIVcpzGAB1. These findings indicate that naturally occurring SIVcpz strains fall into two related yet highly divergent phylogenetic lineages.

To explore whether a host-dependent evolution of SIVcpz could account for the extraordinary diversity between SIVcpzANT and the other three SIVcpz strains, the researchers determined the subspecies identity of the animals from which these viruses were derived. In Fig 2(a), four chimpanzee subspecies with non-overlapping geographic ranges have been proposed on the basis of differences in mitochondrial (mt) DNA sequences [5]. The researchers amplified and sequenced a 498-bp fragment of mitochondrial control region (D-loop) sequences from Peripheral-Blood Mononuclear Cell (PBMC) or spleen DNA of the four SIVcpz-infected chimpanzees. In Fig 2(b), comparison of these newly derived mtDNA sequences to representative sequences from the four chimpanzee subspecies revealed that the three chimpanzees infected with the more closely related SIVcpzGAB1 (GAB1), SIVcpzGAB2 (GAB2) and SIVcpzUS (Marilyn) strains all belonged to the *Pan troglodytes troglodytes*. Thus there has been host-dependent evolution of SIVcpz in chimpanzees.

To look for evidence of cross-species transmission, in Fig 2(c), a comparison is made between the phylogenetic positions of the three major groups (termed M, N and O) of

globally circulating strains of HIV-1 and those of the four SIVcpz strains. The result shows that all three HIV-1 groups (M, N and O) clustered closely only with SIVcpz strains infecting chimpanzees of the *Pan troglodytes troglodytes*. This applied for all coding regions and using different phylogenetic methods. These strongly indicate that HIV-1 infection of humans occurred as a result of cross-species transmissions of SIVcpz from *Pan troglodytes troglodytes*.

There area two additional lines of evidence supported *Pan troglodytes troglodytes* origin of HIV-1 origin. First, YBF30, the only fully sequenced example of HIV-1 group N, is found as a recombinant of divergent viral lineages within the HIV-1/SIVcpz (*Pan troglodytes troglodytes*) group. In Fig 3(a), by distance plots of full-length (concatenated) protein sequences revealed that YBR30 and SIVcpzUS were disproportionately more similar to each other in the 3' half compared to the 5' half of their genome. In Fig 3(b), phylogenetic tree analyses confirmed these discordant relationship, showing that YBF30 fell into significantly different phylogenetic positions in different parts of it genome. This mosaic genome structure of YBF30 implies previous co-infection and recombination of divergent SIVcpz strains in a *Pan troglodytes troglodytes* host. Second, by carefully analyzing three full-length SIVcpz genomes for chimpanzee specific 'signature' sequences, a single protein domain, the V3 loop region of the extracellular envelope glycoprotein is found to be conserved uniquely among all SIVcpz strains. This sequence conservation was most evident in the V3 crown region which was identical among the three chimpanzee viruses and differed by only a single amino-acid residue in YBF30. These data indicate that YBF30, by virtue of its similarity to SIVcpz in V3, may represent a virus lineage most recently transmitted to human.

## III Conclusions

Generally, five lines of evidence have been used to substantiate zoonotic transmission of primate lentiviruses: first, similarities in viral genome organization; second, phylogenetic relatedness; third, prevalence in the natural host; fourth, geographic coincidence; fifth, plausible routes of transmission.

In discussions above, genome similarities and close phylogenetic relationship between HIV-1 stains and SIVcpz strains infecting *Pan troglodytes troglodytes* are clearly demonstrated. Third, the detection of recombination among divergent SIVcpz lineages provides further evidence that SIVcpz infection rates in wild-living chimpanzees must have been (and still may be) substantial; a trivial explanation for the observed low frequency of SIVcpz infection in captive chimpanzees is that such animals were either born in captivity or captured as infants before they matured and had increased risk for SIVcpz infection. Fourth, the natural range of *Pan troglodytes troglodytes* coincides precisely with areas of HIV-1 group M, N and O endemicity. Last, as chimpanzees are commonly hunted for food, especially in west equatorial Africa, thus represents a ready source for zoonotic transmissions of SIVcpz to man. Actually, in Fig 3(b), right panel indicate that the three HIV-1 groups have each arisen as a consequence of independent zoonotic transmissions of SIVcpz from *Pan troglodytes troglodytes* to man.

To conclude, all HIV-1 strains known to infect man, including HIV-1 groups M, N and O, are closely related to just one of these SIVcpz lineages, that found in *Pan troglodytes troglodytes*. Moreover, it is found that HIV-1 group N is a mosaic of SIVcpzUS and HIV-1 related sequences, indicating an ancestral recombination event in a chimpanzee host. There results, together with the observation that the natural range of *Pan troglodytes troglodytes* coincides uniquely with areas of HIB-1 group M, N and O endemicity, indicate that *Pan troglodytes troglodytes* is the primary reservoir for HIV-1 and has been the source of at least three independent introductions of SIVcpz into the human population.

It is still possible, however, that the other chimpanzee subspecies are also infected with SIVcpz had have transmitted their viruses to humans. Such transmissions have not been detected but could have gone unrecognized because of the explosive spread of HIV-1 group M and the absence of serological tests to distinguish SIVcpz (*Pan troglodytes troglodytes*) from other SIVcpz lineages.

From my point of view, although discussions and reasoning above are convincing, some observation and speculation are still not strict enough. For example, chimpanzees are rarely observed to be infected with SIVcpz. Thus, it should be not necessarily clear that chimpanzees are the *original* reservoir for HIV-1; it is still possible that both chimpanzees and humans have been infected from a third, as yet unidentified, primate species.

[1] Gao, F., et al., Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes, *Nature* 397 (1999) 436-441

[2] Heseltine, W. A., Molecular–biology of the human-immunodeficiency-virus type-1, *FASEB Journal* 5 (1990) 2349-2360

[3] Huet, T., et al., Genetic organization of a chimpanzee lentivirus related to HIV-1, *Nature* 345 (1990) 356-359

[4] Hirsch, V. M., et al., An African primate lentivirus (SIVsm) closely related to HIV-2, *Nature* 339 (1989) 389-392

[5] Gonder, M. K. et al., A new west African chimpanzee subspecies, *Nature* 388 (1997) 337

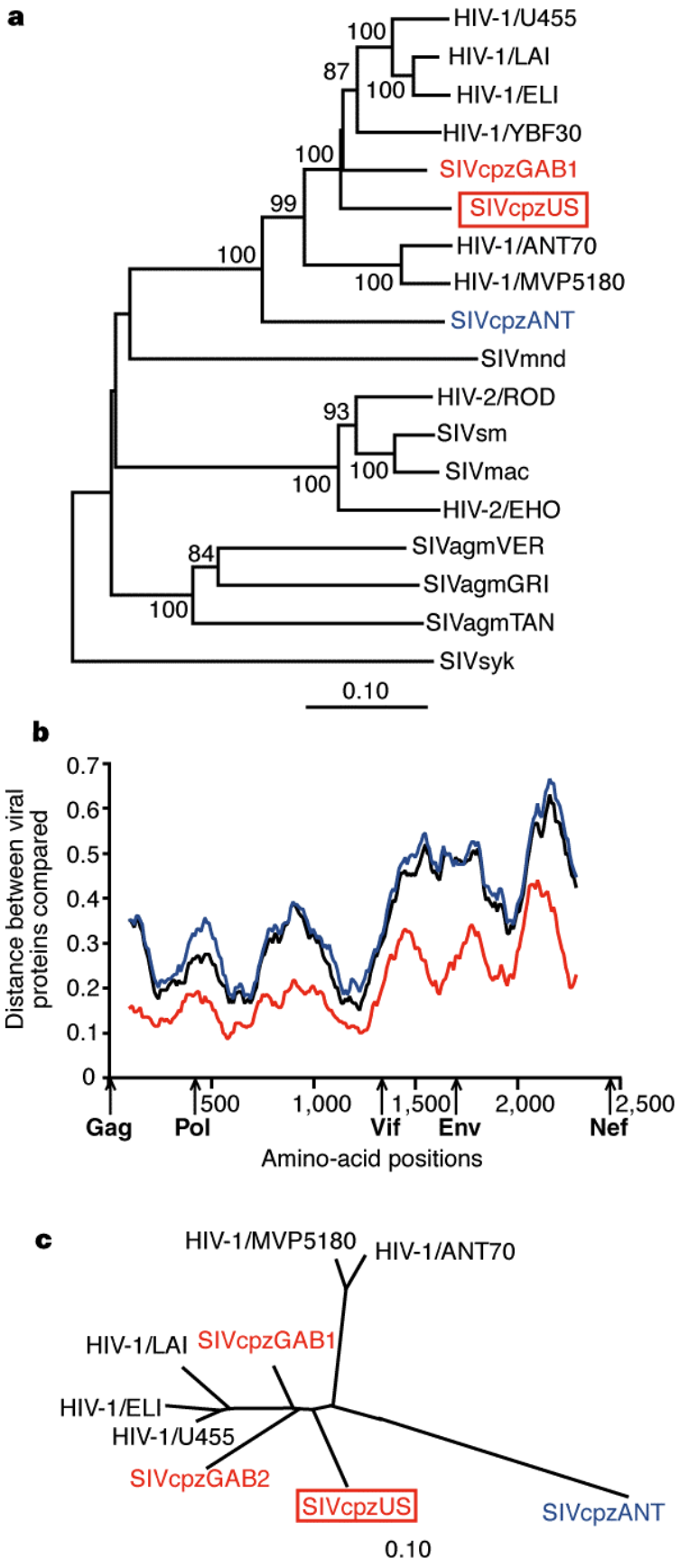
a
HIV-1/U455
HIV-1/LAI
HIV-1/ELI
HIV-1/YBF30
SIVcpzGAB1
SIVcpzUS
HIV-1/ANT70
HIV-1/MVP5180
SIVcpzANT
SIVmnd
HIV-2/ROD
SIVsm
SIVmac
HIV-2/EHO
SIVagmVER
SIVagmGRI
SIVagmTAN
SIVsyk
100
87
100
100
99
100
100
93
100
100
84
100
0.10
b
Distance between viral proteins compared
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
500
1,000
1,500
2,000
2,500
Gag
Pol
Vif
Env
Nef
Amino-acid positions
c
HIV-1/MVP5180
HIV-1/ANT70
HIV-1/LAI
SIVcpzGAB1
HIV-1/ELI
HIV-1/U455
SIVcpzGAB2
SIVcpzUS
SIVcpzANT
0.10

**Figure 1** Phylogenetic analysis of SIVcpzUS. **a**, Phylogenetic relationship of SIVcpzUS to other primate lentiviruses. The tree was derived by neighbour-joining analysis[27] of full-length Pol sequences (trees derived by maximum-likelihood methods[28] yielded very similar topologies). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 amino-acid replacements per site. Numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Other SIVcpz strains closely or more distantly related to SIVcpzUS are shown in red and blue, respectively. **b**, Diversity plots of concatenated SIVcpz protein sequences depicting the proportion of amino-acid sequence differences between SIVcpzUS and SIVcpzGAB1 (red), SIVcpzUS and SIVcpzANT (blue), and SIVcpzGAB1 and SIVcpzANT (black), calculated for a window of 200 amino acids moved in steps of 10 amino acids along the alignment (available as Supplementary Information). The *x*-axis shows the amino-acid positions along the alignment. The positions of Gag, Pol, Vif, Env and Nef regions are shown. The *y*-axis denotes the distance between the viral proteins compared (0.1 = 10% difference). **c**, Unrooted neighbour-joining tree of partial Pol protein sequences (distances are drawn to scale).

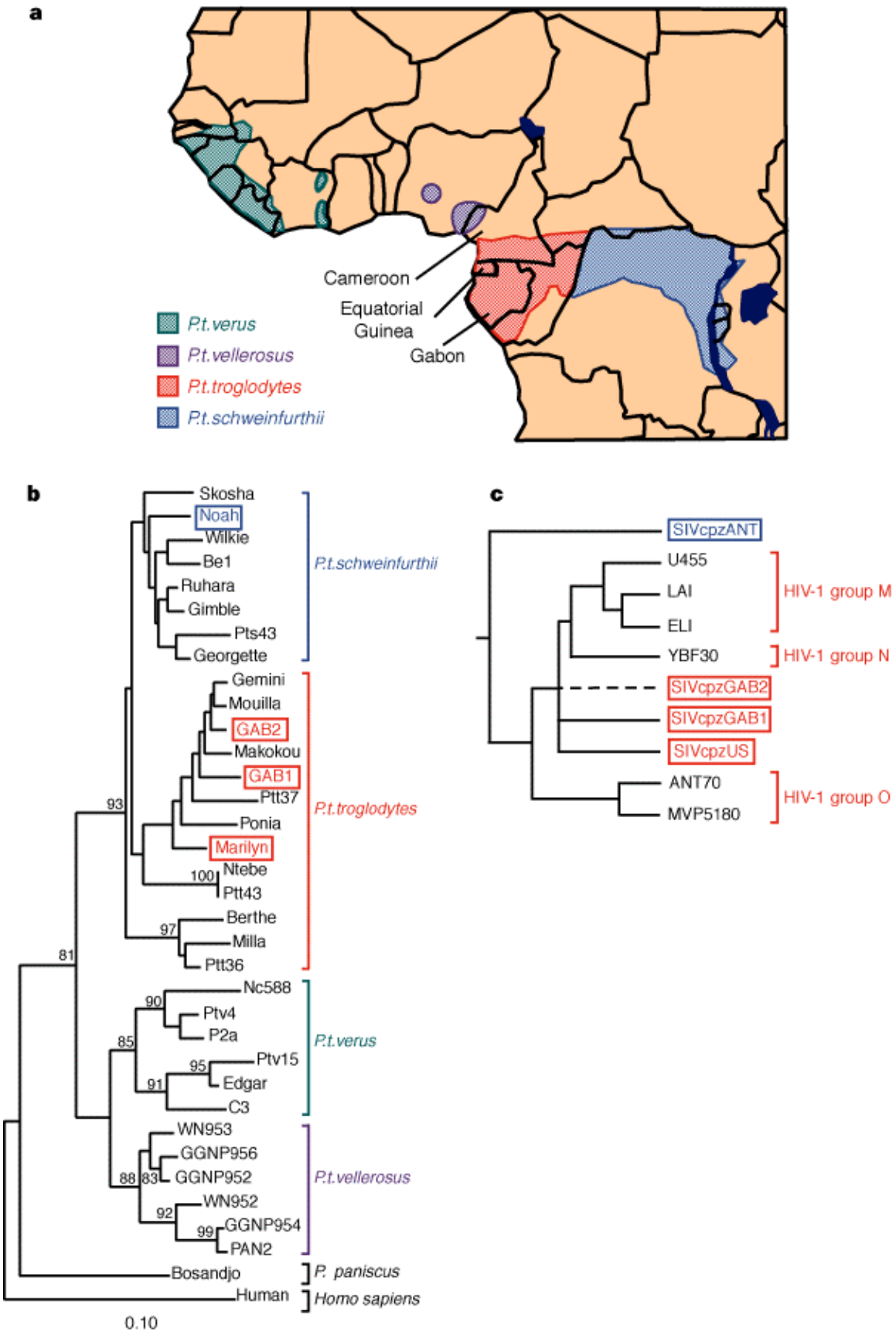
a
Cameroon
Equatorial Guinea
Gabon
P.t.verus
P.t.vellerosus
P.t.troglodytes
P.t.schweinfurthii
b
Skosha
Noah
Wilkie
Be1
Ruhara
Gimble
Pts43
Georgette
Gemini
Mouilla
GAB2
Makokou
GAB1
Ptt37
Ponia
Marilyn
Ntebe
Ptt43
Berthe
Milla
Ptt36
Nc588
Ptv4
P2a
Ptv15
Edgar
C3
WN953
GGNP956
GGNP952
WN952
GGNP954
PAN2
Bosandjo
Human
P.t.schweinfurthii
P.t.troglodytes
P.t.verus
P.t.vellerosus
P. paniscus
Homo sapiens
93
81
90
85
95
91
88
83
92
99
97
100
0.10
c
SIVcpzANT
U455
LAI
ELI
YBF30
SIVcpzGAB2
SIVcpzGAB1
SIVcpzUS
ANT70
MVP5180
HIV-1 group M
HIV-1 group N
HIV-1 group O

**Figure 2** Origin of HIV-1 in *Pan troglodytes troglodytes*. **a**, Geographic ranges of the four subspecies of the common chimpanzee (*Pan troglodytes*) defined by mtDNA analysis (adapted from refs 19, 20 with permission). **b**, Phylogenetic tree of mtDNA sequences. Positions of sequences from the SIVcpz-infected chimpanzees Marilyn (SIVcpzUS), GAB1 (SIVcpzGAB1), GAB2 (SIVcpzGAB2) and Noah (SIVcpzANT) are boxed. The phylogeny was derived by the neighbour-joining method[27] applied to pairwise sequence distances calculated using the Kimura two-parameter method (transition/transversion ratio set to 10). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 nucleotide replacements per site. Numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Brackets on the right indicate previously defined subspecies/species classifications[19, 20] (*P. t. troglodytes* , *P. t. schweinfurthii*, *P.t.verus*, and *P. t. vellerosus* are colour coded as in **a**). **c**, Schematic tree of Pol sequences, highlighting the position of HIV-1 group M, N and O viruses in relation to *P. t. troglodytes* (red) and *P. t. schweinfurthii* (blue) viruses. The position of SIVcpzGAB2 (indicated by broken line), for which only partial sequence is available[10], is inferred from the phylogeny shown in Fig. 1c.

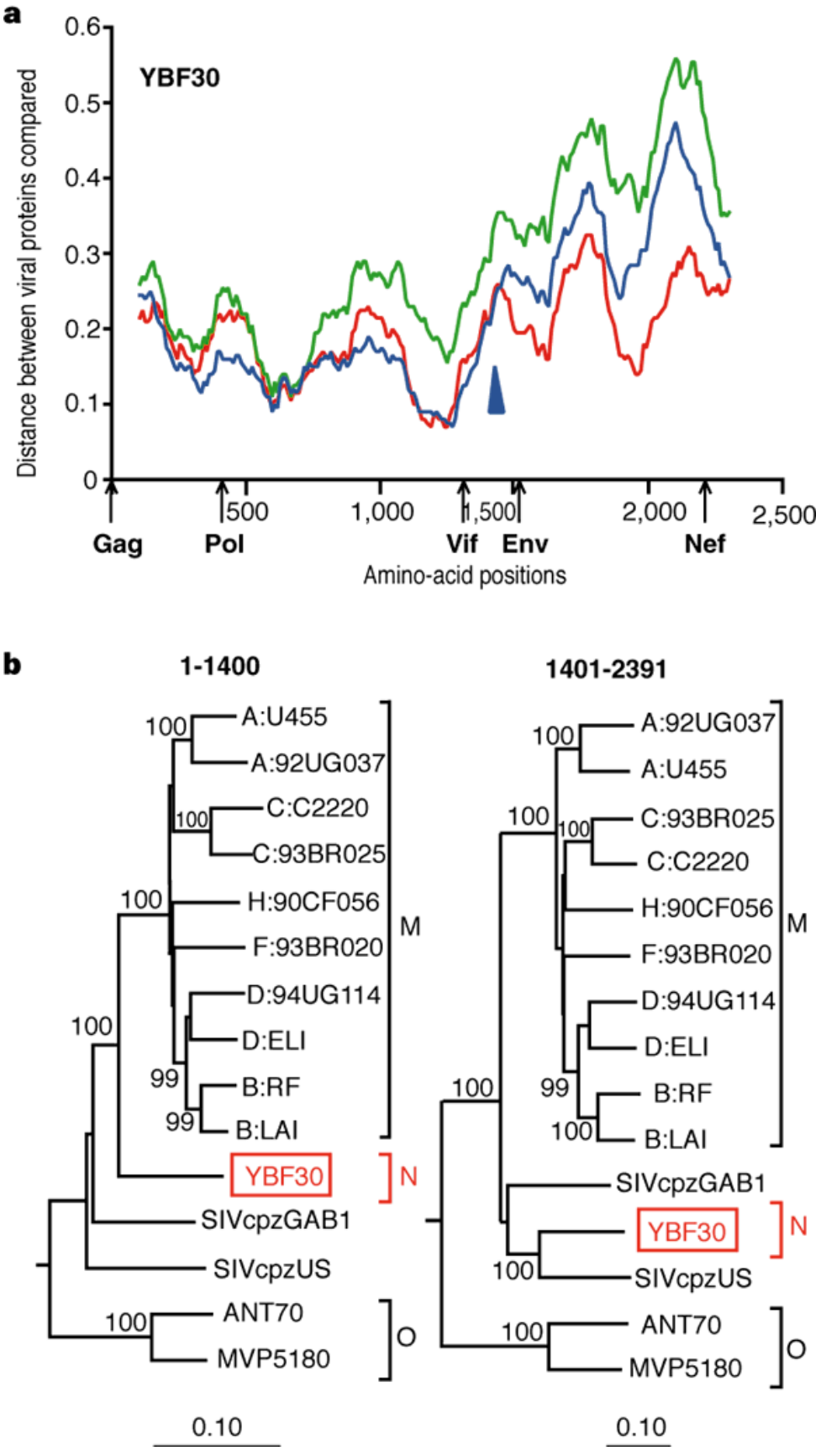
a
YBF30
Distance between viral proteins compared
0.6
0.5
0.4
0.3
0.2
0.1
0
500
1,000
1,500
2,000
2,500
Gag
Pol
Vif
Env
Nef
Amino-acid positions
b
1-1400
1401-2391
A:U455
A:92UG037
C:C2220
C:93BR025
H:90CF056
F:93BR020
D:94UG114
D:ELI
B:RF
B:LAI
M
YBF30
N
SIVcpzGAB1
SIVcpzUS
ANT70
MVP5180
O
100
99
0.10

**Figure 3** Recombinant origin of HIV-1/YBF30 (group N). **a**, Diversity plots of concatenated protein sequences, depicting the proportion of amino-acid sequence differences between YBF30 (HIV-1 group N) and SIVcpzUS (red), U455 (HIV-1 group M; blue), and MVP5180 (HIV-1 group O; green), were calculated for a window of 200 amino acids moved in steps of 10 amino acids along an alignment. The *x*-axis indicates the amino-acid positions along the alignment. The positions of Gag, Pol, Vif, Env and Nef regions are shown. The *y*-axis denotes the distance between the viral proteins compared (0.1 = 10% difference). A blue marker at position 1,400 delineates 5' and 3' regions of disproportionate sequence similarity between YBF30 and SIVcpzUS. **b**, Phylogenetic position of YBF30 (boxed) in different parts of its genome. Trees were derived by neighbour-joining analysis[27] of concatenated protein sequences flanking the putative recombination breakpoint indicated by the marker in **a** (discordant phylogenies for YBF30 were confirmed by maximum-likelihood methods[28]). Horizontal branch lengths are drawn to scale with the bar indicating 0.1 amino-acid replacements per site; numbers at each node indicate the percentage of bootstrap samples (out of 1,000) in which the cluster to the right is supported (only values >80% are shown). Brackets identify members of HIV-1 groups M, N and O.