Originally published as:

**Li, Y., Ndjango, J.-B., Learn, G.H., Ramirez, M.A., Keele, B.F., Bibollet-Ruche, F., Liu, W., Easlick, J.L., Decker, J.M., Rudicell, R.S., Inogwabini, B.-I., Ahuka-Mundeke, S., Leendertz, F.H., Reynolds, V., Muller, M.N., Chancellor, R.L., Rundus, A.S., Simmons, N., Worobey, M., Shaw, G.M., Peeters, M., Sharp, P.M., Hahna, B.H.**
**Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir (2012) Journal of Virology, 86 (19), pp. 10776-10791.**

**DOI: 10.1128/JVI.01498-12**

# Eastern Chimpanzees, but not Bonobos, Represent a Simian Immunodeficiency Virus Reservoir

Yingying Li,[1] Jean-Bosco Ndjango,[2] Gerald H. Learn,[1] Miguel Ramirez[1], Brandon F. Keele,[3] Frederic Bibollet-Ruche,[1] Weimin Liu[1], Juliet L. Easlick,[4] Julie M. Decker,[4] Rebecca S. Rudicell,[4#] Bila-Isia Inogwabini,[5,6] Steve Ahuka-Mundeke,[7,8] Fabian H. Leendertz[9], Vernon Reynolds,[10,11] Martin N. Muller,[12] Rebecca L. Chancellor,[13,14] Aaron S. Rundus,[13,14], Nicole Simmons[15], Michael Worobey,[16] George M. Shaw,[1,17] Martine Peeters,[7] Paul M. Sharp[18] and Beatrice H. Hahn[1,17]

Departments of [1]Medicine and [17]Microbiology, University of Pennsylvania, Philadelphia, PA 19104, USA; [2]Department of Ecology and Management of Plant and Animal Resources, Faculty of Sciences, University of Kisangani, Kisangani, Democratic Republic of the Congo; [3]The AIDS and Cancer Virus Program, Science Applications International Cooperation-Frederick Inc., National Cancer Institute-Frederick, MD 21702, USA; [4]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [5]Lac Tumba Project, World Wildlife Fund, Kinshasa, Democratic Republic of the Congo; [6]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala SE-750 07, Sweden; [7]Institut National de Recherche Biomedicales, Kinshasa, Democratic Republic of the Congo; [8]UM1 233, Institut de Recherche pour le Développement (IRD) and University of Montpellier 1, Montpellier, France; [9]Research Group Emerging Zoonoses, Robert Koch-Institute, Berlin, Germany; [10]School of Anthropology, Oxford University, Oxford OX2 6PE, UK; [11]Budongo Conservation Field Station, Masindi, Uganda; [12]Department of Anthropology, University of New Mexico, Albuquerque, NM 87131, USA; [13]Department of Psychology, West Chester University, West

Chester, PA 19383; [14]Gishwati Area Conservation Program, Great Ape Trust, Gisenyi, Rwanda; [15]Department of Zoology, Makerere University, Kampala, Uganda; [16]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA; [18]Institute of Evolutionary Biology and Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom.

#present address: Vaccine Research Center, National Institute of Health, Bethesda, MD 20892

Running title: Widespread SIVcpz infection in eastern chimpanzees

## ABSTRACT

Chimpanzees in west central Africa (*Pan troglodytes troglodytes*) are endemically infected with simian immunodeficiency viruses (SIVcpz*Ptt*) that have crossed the species barrier to humans and gorillas on at least five occasions, generating pandemic and non-pandemic forms of human immunodeficiency virus type 1 (HIV-1) as well as SIVgor. Chimpanzees in east Africa (*P. t. schweinfurthii*) are also infected with SIVcpz; however, their viruses (SIVcpz*Pts*) have never been found in humans. To examine whether the latter is due to a paucity of natural infections, we used non-invasive methods to screen wild-living eastern chimpanzees in the Democratic Republic of the Congo (DRC), Uganda and Rwanda. We also screened bonobos (*Pan paniscus*) in the DRC, a species not previously tested for SIV in the wild. Fecal samples (n=3,108) were collected at 50 field sites, tested for species and subspecies origin, and screened for SIVcpz antibodies and nucleic acids. Of 2,565 samples from eastern chimpanzees, 323 were antibody positive and 92 contained viral RNA. The antibody positive samples represented 76 individuals from 19 field sites, all sampled north of the Congo River in an area spanning 250,000 $km^2$. In this region, SIVcpz*Pts* was common and widespread, with seven field sites exhibiting infection rates of 30% or greater. The overall prevalence of SIVcpz*Pts* infection was 13.4% (95% CI 10.7% - 16.5%). In contrast, none of 543 bonobo samples from six sites was antibody positive. All newly identified SIVcpz*Pts* strains clustered in strict accordance to their subspecies origin; however, they exhibited considerable genetic diversity, especially in protein domains known to be under strong host selection pressure. Thus, the absence of SIVcpz*Pts* zoonoses cannot be explained by an insufficient primate reservoir. Instead, greater adaptive hurdles may have prevented the successful colonization of humans by *P. t. schweinfurthii* viruses.

## INTRODUCTION

Of over 40 African primate species naturally infected with simian immunodeficiency viruses (SIVs), chimpanzees (*Pan troglodytes*) are unique because they harbor the virus (SIVcpz) that spawned the human AIDS pandemic (20, 30, 78). It is now well established that chimpanzees are the original source of viruses currently found in chimpanzees, gorillas and humans, and that ape viruses have been transmitted to humans on at least four occasions, generating human immunodeficiency virus type 1 (HIV-1) groups M, N, O and P (24, 65, 66). Chimpanzees also differ from most other primate species by having acquired their infection relatively more recently, as a consequence of cross-species transmission and recombination of viruses infecting monkeys on which they prey (3). Importantly, natural history studies have shown that SIVcpz is quite pathogenic and has a substantial negative impact on the health, reproduction and life span of its natural host (17, 29, 55, 75). Thus, in addition to representing a potential source for human infection, SIVcpz also comprises a serious threat to chimpanzee populations living in the wild (55).

Chimpanzees are highly endangered, thus requiring non-invasive approaches to study SIVcpz infection in wild populations (60, 61). To address this, we have -- over the past decade -- developed approaches that permit the detection of virus specific antibodies and nucleic acids in fecal and urine samples (29, 30, 36, 42, 55, 56, 59, 60, 73, 74, 78, 79, 85). Examining the molecular epidemiology of SIVcpz in the wild, we found that only two of the four currently recognized subspecies (9), i.e., central (*P. t. troglodytes*) and eastern (*P. t. schweinfurthii*) chimpanzees, but not western (*P. t. verus*) and Nigeria-Cameroonian *(P. t. ellioti*) chimpanzees, are naturally infected by SIVcpz (30, 35, 56, 59, 60, 78). We also found that SIVcpz is unevenly distributed among wild apes, with high prevalence rates (up to 50%) in some communities, and rare or absent infection in others (29, 30, 55, 56, 60, 78). Molecular studies of existing SIVcpz strains revealed that they cluster in accordance to their subspecies origin in two highly divergent

lineages, termed SIVcpz*Ptt* and SIVcpz*Pts* (66). Interestingly, all groups of HIV-1 as well as viruses (SIVgor) from western gorillas (*Gorilla gorilla gorilla*) fall into just one of these lineages, clustering with SIVcpz*Ptt* from *P. t. troglodytes*, implicating the central subspecies as the source of both human and gorilla infections (30, 66, 73, 79). In contrast, evidence for transmission of SIVcpz*Pts* strains to either humans or sympatric eastern gorillas (*G. g. beringei*) is lacking (42), raising questions as to the relative abundance and distribution of SIVcpz in wild-living *P. t. schweinfurthii* populations (46).

Eastern chimpanzees live in central and east Africa (Fig. 1A), in an area that ranges from the southeastern parts of the Central African Republic (CAR) through the northern parts of the Democratic Republic of Congo (DRC), to southwestern Sudan and the western parts of Uganda, Rwanda, Burundi and Tanzania (51). The first SIVcpz*Pts* strain (ANT) was identified in an ape captured at an unknown location in the DRC and exported to Belgium (45, 80, 20). Subsequent testing of fecal samples from a limited number of chimpanzees in the vicinity of Kisangani indicated that SIVcpz*Pts* was also present in wild-living *P. t. schweinfurthii* communities in the DRC (85). However, extensive field studies in east Africa revealed a surprising paucity of SIVcpz*Pts* infections. Although infected chimpanzees were identified in Gombe National Park (29, 55, 60, 61) and the Ugalla region of western Tanzania (56), communities in the Budongo Forest (BG) in northern Uganda, the Kibale National Park (KB) in western Uganda, the Bwindi Impenetrable Forest (BW) in southern Uganda, and Mahale Mountains National Park (MH) in Tanzania seemed free of SIVcpz infection (56, 60, 67). Moreover, none of over 300 fecal samples from the Nyungwe Forest Reserve (NY) in western Rwanda was virus positive (67). Thus, in contrast to the high infection rates observed for *P. t. troglodytes* apes in southern Cameron (30, 42, 78) and northern Gabon (36), field studies at the eastern limits of the *P. t. schweinfurthii* range identified only isolated foci of infection.

The habitat of eastern chimpanzees in Uganda, Rwanda and Tanzania is severely fragmented (Fig. 1B), due to extensive deforestation, expanding agriculture and human

encroachment (51). To test whether this habitat loss might have contributed to the paucity of SIVcpz*Pts* infection in east Africa, we targeted wild-living apes in the Democratic Republic of the Congo, a country that is home to half of the world's remaining chimpanzees (10, 51). There are two different chimpanzee species in the DRC, the eastern chimpanzee (*Pan troglodytes schweinfurthii*) and the bonobo (*Pan paniscus*), who live in non-overlapping ranges north and south of the Congo River, respectively (Fig. 1A). Current population estimates suggest that there may be as many as 180,000 to 200,000 eastern chimpanzees and 30,000 to 50,000 bonobos still remaining in largely intact forest blocks (10, 19, 51). With the exception of one study (85), wild-living apes in the DRC have not previously been surveyed for SIV. Here, we show that eastern chimpanzees, but not bonobos, are widely and commonly infected with SIVcpz*Pts* and thus represent a substantial virus reservoir.

## MATERIALS AND METHODS

**Study sites and sample collection.** The vast majority of ape fecal samples were collected in the DRC from non-habituated eastern chimpanzees (n=2,480) and bonobos (n=543) by teams of local trackers (Table 1). Samples were collected in the vicinity of chimpanzee night nests or when encountered during forest walks, placed into 50 ml conical tubes, and preserved in an equal volume of RNA*later* (Life Technologies) as described (30, 60, 78). Tubes were labeled with a sample number, the field site code, and GPS coordinates when available. Because local trackers were participating in the collection effort, the quality of samples varied between field sites and individual specimens were frequently divided into multiple aliquots without this being indicated. In four instances, samples were collected from pet chimpanzees kept by local villagers (BU203, KS133, KS134 and KS135). Samples were also collected from habituated chimpanzee communities in the Budongo Reserve (BG) in northern Uganda (n=20) (53) and the Kyambura Gorge (KY) in western Uganda (n=16) (32). At the latter two sites, fecal

samples were collected from individually known apes under direct observation by resident primatologists. Finally, samples were obtained from non-habituated chimpanzees in the Gishwati Forest (GI) in northwestern Rwanda (n=49) (12, 50). Because of a lack of refrigeration at most field sites, especially in the DRC, samples were kept at ambient temperature for varying periods of time (usually several weeks, but in some instances several months) before they could be stored at -20°C. In the DRC, this was done at a central laboratory in Kisangani, where samples were then batched and shipped to the US. Bonobo samples from the ML field site were stored at -80°C at the Institut National de Recherche Biomédicale in Kinshasa and then shipped directly to the University of Montpellier.

**SIVcpz antibody detection.** All fecal samples were screened for the presence of HIV-1 cross-reactive antibodies as previously described (30, 42, 60, 78). Bonobo samples from the ML field site were tested using the INNO-LIA HIV I/II Score Confirmation test (Innogenetics, Ghent, Belgium) (78). All other samples were examined by enhanced chemiluminescent Western immunoblot analysis modified for RNA*later* preserved specimens. RNA*later* is a high salt solution (25 mM sodium citrate, 10 mM EDTA, 70 g ammonium sulfate/100 ml solution, pH 5.2) that preserves nucleic acids, but precipitates proteins, including immunoglobulin. To prepare extracts suitable for Western blot analysis, fecal/RNAlater mixtures (1.5 ml) were diluted with PBS-Tween-20 (8.5 ml), inactivated for 1 hr at 60ºC, clarified by centrifugation (3500 x *g* for 30 min) to remove solid debris, and then dialyzed against PBS overnight at 4ºC to reconstitute fecal immunoglobulin. Reconstituted extracts were subjected to immunoblot analysis using commercially available HIV-1 antigen containing strips (Maxim Biomedical, Inc.). Sample integrity was examined using an IgG control.

**Amplification of SIVcpz virion RNA.** All Western blot positive samples were tested for the presence of SIV nucleic acids by reverse-transcription polymerase chain reaction (RT-PCR)

amplification as described (30, 38, 42, 60, 78). Briefly, fecal RNA was extracted using the RNAqueous Midi Kit (Life Technologies) and subjected to RT-PCR amplification using SIVcpz specific *gag*, *pol*, *vpu*, gp41, and gp41/*nef* consensus primers (Table S1). PCR conditions generally included 60 cycles of denaturation (94 °C, 20 s), annealing (50 °C, 30 s), and elongation (68 °C, 1.5 min) for the first round. Second round conditions included 50 cycles of denaturation (94 °C, 20 s), annealing (52 °C, 30 s), and elongation (68 °C, 1 min). All amplicons were gel purified and sequenced directly. Samples that failed to yield SIVcpz amplicons were re-amplified using pan-SIV specific primers (Table S1), so as to not miss infection with SIVs other than SIVcpz.

**Individual identification.** All fecal samples were subjected to mitochondrial DNA analysis to confirm their species and subspecies origin. In addition, all antibody positive *P. t. schweinfurthii* samples as well as 146 bonobo samples from three different field sites (IK, LK, KR) were subjected to microsatellite analyses. Fecal DNA was extracted as described, and used to amplify a 498 bp fragment of the mitochondrial D-loop region (30, 39, 60). Amplicons were sequenced directly (using one primer which yielded 479 bp of sequence) and classified into distinct mitochondrial haplotypes (Fig. S1; Table S2). To identify the number of sampled individuals, fecal samples were genotyped at four (*P. t. schweinfurthii*) or eight (*P. paniscus*) autosomal microsatellite loci (Tables S3 and S4), with amplification products sized on an automated sequencer using GeneMapper 4.0 (Applied Biosystems). Samples were first grouped by field site and mitochondrial DNA haplotype. Within each haplotype, samples were then grouped by microsatellite genotypes, and when possible also by gender and viral genotype. Due to prolonged storage at ambient temperatures, even mtDNA positive samples were frequently partially degraded. We thus allowed allelic mismatches at up to four (eastern chimpanzees) or six (bonobos) loci, if other markers (mtDNA haplotype, gender) indicated possible identity. This very conservative approach likely resulted in an underestimation of the

number of sampled individuals. Samples with evidence of DNA admixture (multiple peaks for the same locus or double peaks in the mtDNA sequence) were excluded.

**Gender determination**. For most samples, gender was determined by amplifying a 218-bp fragment of the amelogenin gene that contains a 6-bp insertion in the Y, but not the X chromosome, using primers AMEL-F212 (5'-ACCTCATCCTGGGCACCCTGG-3') and AMEL-R212 (5'-AGGCTTGAGGCCAACCATCAG-3') (70). PCR conditions were the same as for the microsatellite analyses and fragments were sized on an automated sequencer. For samples that failed this genotyping, a second set of amelogenin primers (AMXY-1F 5'-CTGATGGTTGGCCTCAAGCCTGTG-3' and AMXY-2R 5'-TAAAGAGATTCATTAACTTGACTG-3') were used to amplify a 977-bp fragment from the X chromosome and a 788-bp fragment from the Y chromosome. The resulting PCR products were sized by 1.2% agarose gel electrophoresis (16, 60).

**SIVcpz prevalence determination**. The prevalence of SIVcpz infection was estimated for each field site based on the proportion of SIVcpz antibody positive fecal samples, but correcting for degradation and redundant sampling. As shown in Table 1, 19% of eastern chimpanzee and 9% bonobo fecal samples failed to yield usable mtDNA sequences and were thus excluded from further analysis. A subset of the remainder was then subjected to microsatellite analyses, which provided a quantitative estimate of oversampling for both species. For *P. t. schweinfurthii*, genotyping of 323 antibody positive fecal samples from 19 field sites identified 76 infected individuals (Table 1), indicating that each had been sampled on average 4.25 times (Table S3). For bonobos, analysis of 146 fecal samples from three field sites identified 72 individuals (Table 1), indicating that each had been sampled on average 2.03 times (Table S4). Using these corrections, the proportion of SIVcpz infected chimpanzees was estimated for each sampling location, taking into account the number of unique mtDNA

217 haplotypes as an indicator of the minimum number of chimpanzees tested. From these
218 determinations, prevalence rates were calculated (95% confidence limits were calculated based
219 on binomial sampling).
220
221 **Full-length genome amplification.** The full-length genome of one representative
222 SIVcpz*Pts* strain from the DRC (BF1167) was sequenced as described (59, 78). Briefly,
223 partially overlapping subgenomic fragments were amplified from fecal RNA, gel purified, and
224 sequenced directly. Chromatograms were examined for positions of base mixtures, and
225 ambiguous sites were resolved as previously reported (7, 56, 73, 74). For positions that did not
226 affect the corresponding amino acid, the predominant nucleotide (highest amplitude in the
227 sequence chromatogram or the most frequent nucleotide in repeat sequencing reactions)
228 was chosen. For positions that affected the corresponding amino acid, the base that encoded
229 the most common amino acid residue in alignments of existing SIVcpz protein sequences was
230 selected. In the absence of an apparent common amino acid (*e.g.*, in hypervariable protein
231 regions), the nucleotide with the highest amplitude in sequence chromatograms was selected.
232 Using these criteria, we were able to infer a unique consensus sequence for BF1167.
233
234 **Construction of a replication competent SIVcpz*Pts* molecular clone.** To obtain a
235 full-length infectious molecular clone of BF1167, the consensus sequence was chemically
236 synthesized as three subgenomic fragments (Blue Heron Biotechnology). These included a 3.7
237 kb 5'-LTR-*pol* fragment, a 3.9 kb *pol*-*env* fragment and a 2.3 kb *env*-3'-LTR fragment. To
238 facilitate subsequent cloning, unique *Mlu*I and *Apa*I sites were added to the 5' and 3' termini of
239 the provirus, respectively. These, together with internal *Nco*I and *Sal*I sites at position 3,728
240 and 7,596 were used to assemble the three subgenomic fragments to generate a full-length
241 provirus. Ligation products were used to transform XL2-MRF bacteria (Stratagene). Resulting
242 transformants were screened for appropriately sized inserts, transfected into 293T cells, and

tested for infectivity in the TZM-bl assay (74) One functional clone (pBF1167) was identified and grown large scale (available from the National Institutes of Health Research and Reference Program, Rockville, MD).

**BF1167 infectivity and co-receptor usage.** BF1167 as well as TAN2 (SIVcpz) and SG3 (HIV-1) reference clones were transfected into 293T cells and supernatants equilibrated by particle-associated RT activity as described (7, 74). Viral infectivity was assessed in TZM-bl cells, a HeLa-derived line which has been genetically-modified to constitutively express human CD4, CCR5 and CXCR4, and to contain integrated luciferase and β-galactosidase reporter genes under the control of an HIV-1 LTR (49, 84). For co-receptor analysis, TZM-bl cells were seeded in 96 well plates at 8,300 cells/well overnight and then treated with the CCR5 antagonist TAK-779 (10μM), the CXCR4 antagonist AMD3100 (1.2μM), or a combination of both for one hour (74). Virus was added in the presence of 40μg/ml DEAE-dextran and removed 48 hours later. Cells were then lysed and analyzed for luciferase activity (Promega) using a Tropix luminometer with WinGlow version 1.24 software.

**CD4 T cell cultures.** Blood was obtained from normal human volunteers as well as healthy (SIV/HIV-1 uninfected) chimpanzees housed at the Yerkes Regional Primate Center as described previously (chimpanzee blood samples were left-over specimens from the annual health check-up) (14). Briefly, peripheral blood mononuclear cells were isolated using Ficoll Hypaque Plus (GE Healthcase). CD4+ T cells were enriched using CD4 microbeads and magnetic cell sorting (Militenyi Biotec), stimulated with staphylococcal enterotoxin B (Sigma-Aldridge) for 12 to 15 hours (3 μg/ml), and subsequently co-cultivated with autologous monocyte derived macrophages for optimal activation (14). After 5 to 6 days, CD4+ T cells were removed from the macrophages, placed into DMEM with 10% FCS, and incubated with 30 U/ml

interleukin-2 (IL-2). After 24 hours, 5 x $10^5$ CD4+ T cells were incubated with transfection-derived viral stocks at a multiplicity of infection (MOI) of 0.1 (as determined on TZM-bl cells) in 300µl DMEM containing 10% FCS and 30 U/ml IL-2 for 16 hours. CD4+ T cells were washed three times, plated in 24-well plates in DMEM with 10% FCS and 30 U/ml IL-2, and reverse transcriptase activity was measured in culture supernatants every three days to monitor viral replication.

**Phylogenetic analyses.** Partial *pol* (232 bp and 892 bp), *vpu/env* (481-514 bp), gp41 (325-465 bp) and gp41/*nef* (699-1259 bp) sequences from the newly characterized viruses were aligned with HIV-1, SIVcpz and SIVgor reference sequences (GenBank accession numbers: HIV-1 group M: HXB2, K03455; HIV-1 group N: YBF30, AJ006022; HIV-1 group O: ANT70, L20587; HIV-1 group P: U14788, HQ179987; SIVcpz*Ptt*: EK505, DQ373065; MB66, DQ373063; LB7, DQ373064 MT145, DQ373066; US, AF103818; CAM13, AY169968; GAB1, X52154; CAM3, AF115393; CAM5, AJ271369; SIVgor, FJ424866; SIVcpz*Pts*: TAN1, AF447763; TAN2, DQ374657; TAN3, DQ373065; TAN5, JN091691 TAN13, JQ768416; UG38, JN091690; and ANT U42720) using CLUSTAL W (34). Regions of the sequences that could not be unambiguously aligned were removed from further analyses. For SIVcpzBF1167, deduced Gag, Pol, Vif, and Env sequences were aligned with the corresponding protein sequences of the same HIV-1, SIVcpz and SIVgor reference strains. Gag/Pol and Pol/Vif protein overlaps were removed from the N- and C-termini of the deduced Pol protein sequences. In addition, the concatenated Pol and Vif alignment was divided into two regions around a previously reported recombination breakpoint (20, 68). Appropriate evolutionary models for phylogenetic analyses were selected using ModelTest version 3.7 (52) and ProtTest version 2.4 (1). For nucleotide sequence analyses these were K80+G for the diagnostic 232 bp *pol* fragment and GTR+I+G for the longer *pol*, *vpu/env*, gp41 and gp41/*nef* fragments. For amino acid sequence analyses the chosen models were LG+I+G (Gag), RtREV+I+G+F (Pol), LG+I+G+F (Pol/Vif) and WAG+I+G+F

(Env). Phylogenetic trees were constructed using maximum likelihood (22) and Bayesian (54) methods, the latter with a 25% burn in and using as a convergence criterion an average standard deviation of partition frequencies < 0.01.

For species and subspecies analysis, mtDNA control region (D loop) sequences were aligned and identical sequences grouped into haplotypes (Table S2). The evolutionary relationships of the new haplotypes to each other and references sequences from the database were then determined by phylogenetic analysis. A neighbor-joining tree is shown in Fig. S1.

**GenBank accession numbers.** New SIVcpz*Pts* sequences have been deposited in GenBank under accession numbers JQ866001, JQ866003-JQ866011, JQ866013-JQ866017, JQ866024, JQ866026-JQ866041, JQ866043-JQ866045, JQ866047-JQ866052, JQ866055, JQ866057-JQ866059, JQ866061-JQ866066, JQ866068-JQ866070 JX178444-JX178449; the new *P. t. schweinfurthii* and *Pan paniscus* mitochondrial D-loop sequences are listed in Table S2 (accession numbers JQ866072-JQ866157, JQ866159-JQ866296).

## RESULTS

**SIVcpz infection is endemic and widespread among eastern chimpanzees.** To determine the geographic distribution and prevalence of SIVcpz in Democratic Republic of the Congo (DRC), we conducted a comprehensive, non-invasive (fecal based) survey at 41 different collection sites (Fig. 1B). Between January 2001 and May 2011, we obtained a total of 2,480 fecal samples north of the Congo River in an area spanning almost 250,000 $km^2$. Given the vastness of this study area, most samples were collected by local trackers in the vicinity of their villages. While this led to some variation in sample quality, we were able to obtain samples as far south as Kasese (KE), as far east as Gombari (GO) and Walikale (WK), as far north as

Bondo (BD) and Niangara (NI), and as far west as Kotakoli (KO) and Bumba (BU) (Fig. 1B). Except for four samples from pet chimpanzees, all other specimens were obtained from wild-living apes within their natural habitat (Table 1). We also obtained samples from habituated *P. t. schweinfurthii* apes in the Budongo Forest (BG) (53) and the Kyambura Gorge (KY) in Uganda, and from a non-habituated group in the Gishwati Forest (GI) in Rwanda (50). The rationale for these surveys was to examine whether previous studies had missed isolated foci of SIVcpz infection in east Africa. Since the latter field sites were well-established, sample collection and storage occurred under more controlled conditions.

Before testing for SIVcpz antibodies and nucleic acids, all specimens were subjected to mitochondrial (mt) DNA analysis to confirm their species origin and assess their integrity. This analysis identified 495 samples that were either not of chimpanzee origin, represented fecal mixtures from more than one individual, or were too degraded for further analysis (Table 1). The remaining 2,070 samples yielded mtDNA (D loop) sequences, which comprised 252 different haplotypes (Fig. S1). All of these were subjected to enhanced chemiluminescence (ECL) Western blot analysis, a method that detects SIVcpz specific antibodies in fecal extracts with high sensitivity (0.92) and specificity (1.00) even after prolonged storage at ambient temperatures (30). Consistent with previous findings, none of the samples collected in Uganda and Rwanda were SIVcpz antibody positive (Table 1). Interestingly, however, 323 fecal specimens from 19 different sites in the DRC exhibited clear evidence of SIVcpz infection (Fig. 2; Table 1). All of these reacted strongly with the HIV-1 p24 core antigen, and 11%, 62% and 35% also reacted with p17 Gag, reverse transcriptase (p66/p55) and integrase (p31) proteins, respectively. Surprisingly, half of the samples (54%) also exhibited cross-reactivity with the HIV-1 envelope antigens (gp41, gp120, gp160), which in some cases was as strong as the human positive control (Fig. 2).

To determine the number of SIVcpz infected apes at the 19 sampling locations, we subjected all antibody-positive samples to microsatellite analyses (Table S3). Many of these

failed to yield a complete genotype, due to partial sample degradation. To guard against allelic drop out, we thus allowed mismatches at up to four microsatellite loci for samples that shared the same mtDNA haplotype. This conservative approach identified a total of 76 SIVcpz infected apes as a minimum estimate (Table 1). Microsatellite analysis also provided a quantitative measure of redundant sampling, which together with the proportion of degraded specimens allowed us to calculate the prevalence of SIVcpz infection at each collection site. As shown in Table 1, analysis of an estimated 567 eastern chimpanzees yielded an overall prevalence rate of 13.4% (95% confidence interval: 10.7% - 16.5%). As previously observed in Cameroon and Tanzania, infection rates at individual field sites varied widely (Table 1), with high prevalence rates observed in some communities and low level or absence of infection in others. In general, evidence for infection was observed throughout the study area, indicating that SIVcpz was not restricted to any one geographic region. The only exceptions were field sites in the northern part of the range, near the Uele River (BT, BD, AN, NI, DL, MA, PO, RU, IS), which seemed to be free of SIVcpz infection. However, the number of usable samples from these sites was very small (n=77), representing only a minor fraction (3.7%) of the total survey. Given this and the extent of sample degradation, it is thus possible that infected communities in this northern area were missed. Seven sites exhibited prevalence rates of 30% or higher, including in the far western (UB) and eastern (MU) parts of the DRC. These results indicate that SIVcpz is common and widespread among most *P. t. schweinfurthii* communities in the DRC (Fig. 1B), with prevalence rates similar to, or exceeding, those previously observed in *P. t. troglodytes* apes in Gabon and Cameroon (30, 36, 42, 78).

**No evidence of SIVcpz infection in wild-living bonobos**. In addition to eastern chimpanzees, the DRC is also home to bonobos (*Pan paniscus*), whose habitat is restricted to forest areas south of the Congo River (19). Although a limited number of bonobos has previously been tested in captivity (77), wild-living members of this species have never been

372 screened for SIVcpz infection. To examine whether bonobos harbor SIV, we collected 543 fecal
373 samples from six field sites located throughout the species range (Fig. 1B). All of these were
374 subjected to mtDNA analysis, which identified 48 degraded or misidentified samples. The
375 remaining 495 samples yielded mtDNA (D loop) sequences that grouped into 24 distinct
376 haplotypes (Fig. S1). These were tested by immunoblot analysis, which failed to detect SIV/HIV
377 cross-reactive antibodies. Western blots were completely negative for all samples, lacking even
378 faint reactivity with the Gag p24 antigen, which is usually the most cross-reactive protein (Fig. 2).
379 To estimate the number of sampled individuals, specimens from three field sites (IK, KR and
380 LK) were genotyped at eight microsatellite loci (Table S4). This analysis identified a minimum of
381 72 individuals and an oversampling factor of 2.03. From this and the proportion of degraded
382 samples, we estimated to have screened approximately 244 individuals, none of whom was
383 SIVcpz infected.
384
385 **SIVcpz*Pts* strains form a monophyletic clade.** To examine the genetic relationships
386 of the *P. t. schweinfurthii* viruses from the DRC to each other and previously characterized
387 SIVcpz strains, all antibody positive samples were subjected to fecal RNA extraction and
388 reverse transcription polymerase chain reaction (RT-PCR) amplification. Using a diagnostic
389 (mini*pol*) primer set (Table S1), we amplified viral sequences from 75 antibody positive
390 specimens. Subsequent screening with gp41 and gp41/*nef* primers identified SIVcpz virion
391 RNA in 17 additional samples, molecularly confirming infection in 25 of the 76 infected
392 individuals (Table S3). So as to not miss infection by other SIVs, we tested the remaining
393 antibody positive samples using pan-SIV specific primers (Table S1). No other sequences were
394 amplified, indicating that the relatively low RT-PCR positivity rate (29%) was the result of partial
395 sample degradation and not infection by other primate lentiviruses. All amplicons were
396 sequenced and phylogenetically analyzed. Although the mini*pol* sequences were relatively short
397 (232 bp), they were of sufficient length to show that all of the new DRC viruses were members

398 of the SIVcpz*Pts* lineage. As shown in Fig. 3, mini*pol* sequences from 20 different individuals
399 grouped with previously characterized *P. t. schweinfurthii* viruses in a clade supported by
400 significant bootstrap values.

401 To examine further the phylogenetic relationships of the newly derived SIVcpz*Pts* strains,
402 we targeted regions in *pol* (892 bp), *vpu-env* (419 bp), gp41 (405 bp) and gp41/*nef* (665 bp) for
403 additional amplifications. Although longer sequences could only be amplified from 18 individuals,
404 their phylogeny confirmed the mini*pol* results (Fig. 4). All newly derived sequences clustered
405 according to their subspecies origin, forming a single well-supported viral lineage. In addition,
406 the new SIVcpz*Pts* strains exhibited evolutionary relationships similar to those previously
407 described for SIVcpz*Ptt* strains (30, 42, 78): viruses from distant collection sites generally
408 formed well-separated clades or lineages, while viruses from the same geographic locale were
409 usually closely related. For example, viruses from Lubutu (LU) and Mungbere (MU) each
410 formed discrete clusters, indicating local transmissions. Significant clustering was also
411 observed for viruses from Bongbola (BL) and Mongandjo (MO), suggesting unimpeded virus
412 flow between these neighboring sites. Interestingly, the BL and MO strains were also closely
413 related to SIVcpzANT, suggesting a possible geographic origin for this reference strain.
414 Nonetheless, phylogeographic clustering was not uniform. Ape communities at Kabuka (KA)
415 and Parisi (PA) harbored multiple divergent SIVcpz*Pts* strains, perhaps indicating a greater
416 connectivity of these communities. There was also evidence of recombination, as would be
417 expected in populations co-infected with divergent viral lineages. One strain from Parisi (PA1)
418 had a mosaic genome as evidenced by its discordant clustering in *pol* and gp41/nef regions (Fig.
419 4A and D). Importantly, however, there was no evidence of recombination between any of the
420 newly identified DRC viruses and SIVcpz*Ptt* strains. In fact, none of the DRC viruses, including
421 those identified at the western most collection sites (UB, BL, MO), were particularly closely
422 related to *P. t. troglodytes* viruses. Overall, SIVcpz*Pts* strains were quite diverse, with

nucleotide sequence distances of up to 25% and 35% in *po*l and gp41/*nef* regions, respectively, compared to 22% and 33% for members of the SIVcpz*Ptt* group.

To obtain at least one full-length SIVcpz*Pts* sequence from the DRC, we selected a sample (BF1167) with a sufficiently high viral load for whole genome amplification. Using strain specific primers, we amplified 12 partially overlapping fragments, which together comprised a complete proviral genome (Fig. 5A). Inspection of the BF1167 consensus sequence revealed uninterrupted open reading frames for all structural and regulatory proteins as well as intact regulatory elements. BF1167 also contained all previously identified SIVcpz*Pts* signatures (56, 59, 74), including three amino acid insertions in Gag p24 and a conserved PPLP Vif motif, a short Vpr protein of 95 amino acids, a 5 amino acid deletion at the C-terminus of Nef, and an insertion in the ectodomain of gp41. Phylogenetic analysis of full-length Gag, Pol and Env proteins showed that BF1167 fell within the SIVcpz*Pts* radiation (Fig. 6), confirming the relationships derived from the partial genome sequences.

**Generation of a replication competent SIVcpz*Pts* clone**. The genomic organization of BF1167 suggested that it may encode a replication competent provirus. To test this, we synthesized its consensus sequence as three subgenomic fragments and ligated them into a low copy number vector (Fig. 5A). The resulting plasmid clone was transfected into 293T cells and culture supernatant used to infect CD4+ T cell cultures. As shown in Fig. 5B, BF1167 derived virus replicated efficiently and to high titers in CD4+ T cells from select human (n=4) and chimpanzee (n=4) donors, with kinetics similar to previously characterized SIVcpz*Pts* (TAN2) and HIV-1 (SG3) strains. Testing its coreceptor usage, we found that the infectivity of BF1167 was completely blocked by the CCR5 antagonist TAK-779, but not by the CXCR4 antagonist AMD3100 (Fig. 5C). This was also true for R5-tropic reference strains of HIV-1 (YU-2), SIVcpz*Pts* (TAN2) and SIVcpz*Ptt* (MB897), but not for X4 (NL4-3) and R5/X4 dual tropic (WEAU) controls (Fig. 5C). Taken together, these data indicate that the newly derived BF1167

449 clone encodes an R5 tropic SIVcpz*Pts* strain capable of infecting both primary human and
450 chimpanzee CD4+ T cells.
451
452 **SIVcpz*Pts* strains require more mutational steps than SIVcpz*Ptt* strains to gain**
453 **human-specific Gag matrix and Vpu adaptations.** To evaluate the zoonotic potential of the
454 newly derived DRC viruses, we compared their sequences in protein domains known to be
455 under strong host specific selection pressure. One such site was previously mapped to position
456 30 (Gag-30) of the viral matrix protein (83). Inspection of the proteome of all available SIVcpz
457 and SIVgor strains identified a Met or Leu at this position. However, when these viruses infected
458 humans, this residue was changed to an Arg in the inferred ancestors of HIV-1 groups M, N,
459 and O, and subsequently to a Lys in some M, N and O strains (83). To determine the nature of
460 Gag-30 in the DRC viruses, we amplified and sequenced the corresponding *gag* fragment from
461 seven strains (Table S3). Interestingly, we found that two of the new viruses (MO1 and UB6)
462 encoded a Met at Gag-30, similar to SIVcpzANT and all known SIVcpz*Ptt* and SIVgor strains
463 (Fig. 7). However, the other five DRC strains encoded a Leu at Gag-30, similar to SIVcpz*Pts*
464 strains from Gombe (TAN) and Ugalla (UG). We then counted how many nucleotide
465 substitutions would be required to change these Gag-30 codons to the HIV-1 specific residues.
466 Arg is encoded by either CGN (N = A, C, T or G) or AGR (R = A or G), and Lys is encoded by
467 AAR. Thus, changing a Met (ATG) to Arg or Lys requires only a single nucleotide substitution.
468 In contrast, Leu is encoded by CTN or TTR, and changing a Leu to Arg or Lys can therefore
469 require one to three nucleotide substitutions (Fig. 7). Examining the Gag-30 codon in all known
470 SIVcpz*Pts* strains, we found that 10 of 14 viruses, including four of seven DRC strains, encoded
471 Leu using TTA or TTG codons and thus required at least two mutational steps to acquire the
472 basic Arg or Lys residues at Gag-30 (Fig. 7).
473 Human specific adaptation has also shaped the function of the HIV-1 Vpu protein (18, 37,
474 62). Vpu modulates the cell surface expression of a number of immunoregulatory proteins,

including the CD4 receptor, the natural killer (NK) cell ligand NTB-A and the lipid antigen presenting protein CD1d (13, 41, 64). In HIV-1, Vpu also counteracts tetherin, an innate restriction factor that inhibits the release of nascent virus particles from infected cells by "tethering" them to the cell surface (43, 47, 76). In SIVcpz and SIVgor, Vpu lacks this function and these viruses antagonize tetherin via their Nef proteins (62). However, these same Nef proteins are inactive against human tetherin due to a five amino acid deletion that confers resistance (62). To gauge how difficult or easy it would be for the newly characterized DRC strains to acquire anti-human tetherin activity, we aligned their Vpu sequences, as well as those of available SIVcpz*Ptt* or SIVgor strains, to the HIV-1 group M Vpu consensus sequence (Fig. 8). Focusing in particular on the N-terminal transmembrane domain (TMD), which has been shown to interact directly with tetherin via four amino acid residues on the same face of its membrane spanning helix (69), we counted the number of nucleotide substitutions that would be required to gain a functional A/GxxxAxxxAxxxW motif (15, 31, 69, 81, 82). The results ranged from a single substitution for a subset of SIVcpz*Ptt* strains to nine changes for the SIVcpz*Pts* strain ANT, with the closest chimpanzee relatives of HIV-1 groups M (MB879, LB715) and N (EK505) requiring only one or two substitutions, respectively (Fig. 8). Although the adaptive distance depended on the particular virus, SIVcpz*Ptt* and SIVgor strains typically required fewer changes (median 4; range 1-8) than SIVcpz*Pts* strains (median 7; range 3-9) to gain the helix-helix interaction motif, suggesting that they might be more prone to human adaptation. We also examined the cytoplasmic domains of the various Vpus for presence or absence of functional motifs that have previously been shown to play a role in tetherin trafficking and/or degradation, including a YxxΦ motif (57), a DSGxxS β–TrCP binding site (40) and a putative ExxxLV trafficking signal (33). Again, all of these domains were more commonly found in the Vpus of SIVcpz*Ptt* and SIVgor strains than in the Vpus of SIVcpz*Pts* strains (Fig. 8), with only one of 16 *P. t. schweinfurthii* viruses containing the ExxxLV motif that was recently shown to be required for efficient cell-free virion release from CD4 T cells (33). Thus, as a group, SIVcpz*Pts* strains seem to require a

much greater number of mutational steps to gain human specific adaptations than SIVcpz*Ptt* and SIVgor strains.

## DISCUSSION

Although long known to harbor SIVcpz in the wild (60, 61), wild-living eastern chimpanzees have not been thought to represent a virus reservoir. This is because previous field studies in Uganda, Rwanda and Tanzania failed to uncover infected apes at most locations, except for communities in Gombe National Park and the Masito-Ugalla region of western Tanzania (29, 55, 56, 60, 61, 67). However, since these areas of east Africa comprise only a small part of the *P. t. schweinfurthii* range (Fig. 1), we reasoned that the observed paucity of infections might not be representative of the entire subspecies. To test this hypothesis, we targeted wild-living ape populations in the DRC, a country believed to be home to as many as 200,000 *P. t. schweinfurthii* apes and 50,000 bonobos (10, 19). To cover this vast area, we recruited local teams of trackers to collect ape fecal samples in the vicinity of their villages. Although this resulted at times in prolonged sample storage at ambient temperatures and thus partial specimen degradation, we were able to procure an unprecedented number of specimens from a wide variety of different locales (Table 1). This allowed us to assess the prevalence and geographic distribution of SIVcpz throughout the Congo Basin and to determine whether wild-living bonobos, which have not previously been tested for SIV infection, are natural carriers of this virus. We also screened additional communities in Uganda and Rwanda to determine whether isolated foci of SIVcpz infection had previously been missed.

Testing 2,070 fecal samples from an estimated 567 *P. t. schweinfurthii* apes, we identified 323 specimens from 76 individuals to harbor SIVcpz specific antibodies, yielding an overall prevalence of 13.4% (95% confidence interval 10.7% - 16.5%). Since chimpanzees in the northern DRC differ from their east African counterparts in a number of key behavioral traits

(such as ground nesting and lack of termite fishing), which suggests a longstanding cultural separation (25), we also analyzed the DRC populations separately. Excluding samples from Rwanda and Uganda, we found that the remaining sites exhibited a prevalence of 14.9%. Remarkably, this estimate is 2.5 times higher than the prevalence previously determined for *P. t. troglodytes* ape in Cameroon (42). In the latter study, analysis of 1,217 fecal samples from 25 different field sites south of the Sanaga River yielded an overall prevalence of 5.9% (95% confidence interval 4.3% - 7.9%). Although infected apes were identified at 10 of the 25 field sites, local infection levels were relatively low (42). Only one Cameroonian site exhibited a prevalence of greater than 30% (42), compared to seven such sites in the DRC (Table 1). Since the methods of fecal collection and diagnosing SIVcpz were very similar, the observed differences cannot to be explained by ascertainment biases or technical differences. Moreover, infection rates in the DRC represent minimum estimates given that many samples were partially degraded. Although SIVcpz is generally unevenly distributed among wild-living chimpanzees (29, 36, 56, 60), our data indicate that *P. t. schweinfurthii* apes are at least as widely and commonly infected as *P. t. troglodytes* apes. A recent study of wild-living *P. t. troglodytes* apes in Gabon confirmed this, identifying SIVcpz infection at only three of ten locations, with high level infection detected at only one of these sites (36). In contrast, the newly screened communities in the Budongo Reserve (BG), the Kyambura Gorge (KY) and the Gishwati Forest (GI) were all virus negative, further supporting the notion that SIVcpz is absent from the extreme eastern edge of the *P. t. schweinfurthii* range, except for isolated communities in western Tanzania. It remains unknown whether SIVcpz was once present there and has subsequently gone extinct, or whether its eastward spread was obstructed by habitat loss and/or other barriers, such that certain communities were never exposed.

In contrast to eastern chimpanzees, none of 495 bonobo samples from an estimated 244 individuals were SIV antibody positive (Table 1). Although bonobos were sampled at fewer locations than *P. t. schweinfurthii* apes, the six field sites were widely distributed throughout the

553 bonobo range (Fig. 1B). Moreover, for all but one field site (BJ) significant numbers of samples
554 were tested (Table 1), arguing against the possibility that low level infections were missed.
555 There were also no differences in the way bonobo samples were collected, stored and
556 transported compared to those from eastern chimpanzees. Thus, sample degradation cannot
557 be invoked as an explanation for the negative Western blot results. Based on these data, it
558 seems likely that bonobos are free of SIVcpz infection. This is consistent with a previous study
559 that failed to detect SIVcpz antibodies in the blood of 26 captive bonobos (77). It is also
560 consistent with the fact that bonobos and eastern chimpanzees have non-overlapping habitats
561 (Fig. 1). However, given that western gorillas are only rarely infected with SIVgor (42), it will be
562 important to exclude isolated infections of bonobos by testing additional individuals from a wider
563 range of locations. It should also be noted that bonobos, like chimpanzees, are exposed to a
564 variety of SIVs because they hunt and eat smaller primates, including different species of
565 guenons (71, 72), which carry their own types of SIV (2). These monkey SIVs are genetically
566 quite divergent from HIV-1/SIVcpz and may elicit antibodies that do not cross-react with HIV-1
567 proteins (2, 44). It will thus be important to include additional SIV antigens into future non-
568 invasive screening approaches to examine whether bonobos harbor such viruses.

569 Previous evolutionary studies have shown that SIVcpz sequences form two highly
570 divergent lineages, SIVcpz*Ptt* and SIVcpz*Pts,* according to their subspecies of origin (65-67)
571 These studies also revealed that all groups of HIV-1 as well as SIVgor cluster within the
572 SIVcpz*Ptt* lineage, thus identifying *P. t. troglodytes* apes as the original source of both human
573 and gorilla viruses (24, 48, 73, 79). However, except for ANT (80) and the only partially
574 characterized DRC1 virus (85), all other SIVcpz*Pts* strains were derived from infected apes in
575 Gombe and Ugalla (29, 55, 56), thus leaving the evolutionary history of viruses from the
576 remaining parts of the *P. t. schweinfurthii* range open to question (46). In this study, we tested
577 all antibody positive fecal samples from the DRC for the presence of SIVcpz sequences. Using
578 *gag, pol, vpu* and *env* primers, we were able to amplify virion RNA from 25 of the 76 infected

579 individuals (Table S3). Although this recovery rate was lower than in previous field studies, we
580 obtained SIVcpz sequences from 14 of the 19 positive locations (Fig. 1B). Importantly, we
581 found that all of the new viruses clustered with previously characterized SIVcpz*Pts* strains in all
582 genomic regions analyzed (Figs. 3, 4 and 6). While some grouped according to their field site of
583 origin, phylogeographic clustering seemed less pronounced in the DRC than previously
584 observed in Cameroon (30, 78). This may be because the largely contiguous forests in the
585 DRC have provided for greater connectivity and thus exchange of divergent viruses over longer
586 distances. Nonetheless, all DRC viruses fell within the SIVcpz*Pts* radiation, forming a single
587 well supported phylogenetic lineage that also included viruses from Gombe and Ugalla (Figs. 3,
588 4 and 6). This strict subspecies specific clustering indicates that *P. t. schweinfurthii* apes have
589 been effectively isolated from *P. t. troglodytes* apes for a considerable period of time. It also
590 confirms that chimpanzees in the DRC were not the source of any known strain of HIV-1 (85).
591 Given that SIVcpz infection rates in wild-living chimpanzees in the DRC are at least as
592 high, if not higher, than in Cameroon and Gabon, it seems striking that SIVcpz*Pts* strains have
593 never been found in humans. There are at least three potential explanations, which are not
594 mutually exclusive: One possibility is that humans in the DRC are less frequently exposed to
595 SIVcpz. Although it remains unknown exactly how humans acquired the ape precursors of HIV-
596 1 groups M, N, O and P, cross-species transmission must have occurred through cutaneous or
597 mucous membrane exposure to infected ape blood and/or body fluids, which occurs most
598 frequently in the context of bushmeat hunting (24). While firm data concerning the frequencies
599 and types of human-ape interactions in the DRC are lacking, it is believed that hunting varies
600 regionally due to differences in local traditions and preferences, with some tribes having taboos
601 against the consumption of apes (46). Thus, human exposure to SIVcpz may have historically
602 been lower in the DRC compared to west central Africa and this may explain the lack of
603 SIVcpz*Pts* zoonoses. However, even if this were the case, this barrier is clearly no longer in
604 place. Bushmeat hunting in the DRC, including the poaching of chimpanzees, has been on the

605 rise in the past decade due to political turmoil and economic changes (5). A recent study of
606 apes north and south of the Uele River documented a major increase in chimpanzee killing due
607 to an influx of artisanal diamond and gold miners (26). Thus, increased surveillance of humans
608 in these areas for SIVcpz*Pts* and other ape-derived infections may be warranted (6, 11, 39).

609 A second explanation is that SIVcpz*Pts* infections of humans may in fact have occurred,
610 but gone unrecognized, because of limited human sampling and a lack of specific tests. The
611 great majority of HIV-1 infections in the DRC and elsewhere are diagnosed serologically, using
612 enzyme-linked immunosorbent assays (ELISA) or rapid test immunoblot approaches. Despite
613 their genetic diversity, many of the ape infections characterized here exhibited a Western blot
614 profile indistinguishable from that of the positive HIV-1 control (Fig. 2). It is thus possible that a
615 human infected with such a virus could be misdiagnosed as being infected by HIV-1 group M,
616 especially since the vast majority of HIV-1 infections in the DRC are not molecularly confirmed.
617 However, such SIVcpz*Pts* zoonoses -- if they have indeed occurred -- are unlikely to have
618 infected large numbers of people, because more substantive outbreaks would likely have been
619 detected by existing surveillance programs, such as those that discovered the very rare N and P
620 groups of HIV-1 (48, 68). The apparent lack of SIVcpz*Pts* zoonoses could also reflect regional
621 differences in human transmission networks, since sporadic introductions into less dense, less
622 urban and/or less well connected populations would be more likely to result in dead-end
623 infections.

624 A third possibility is that SIVcpz*Pts* strains face greater adaptive hurdles before they can
625 replicate and spread efficiently in the human host. Examining SIVcpz proteomes for amino
626 acids that were highly conserved in the ape precursors of HIV-1, but changed each time these
627 viruses crossed the species barrier to humans, we previously found a Met or Leu in all
628 SIVcpz/SIVgor strains, but an Arg in the inferred ancestors of HIV-1 groups M, N, and O, and a
629 Lys in some HIV-1 strains (83). We also showed that changing Met/Leu at Gag-30 to Arg/Lys
630 greatly enhanced the replication fitness of SIVcpz strains in human tonsil cultures, while the

631 opposite was true for HIV-1 strains that contained the ape-specific Gag-30 residues (7). These
632 studies provided compelling evidence that host specific adaptation at Gag-30 is required for
633 efficient replication of SIVcpz in human lymphatic tissue. When we determined the nature of the
634 Gag-30 residue in all available SIVcpz*Pts* strains, including the new viruses from the DRC, we
635 found that most require twice as many mutational steps than SIVcpz*Ptt* and SIVgor strains to
636 adapt at Gag-30 (Fig. 7). Substantial adaptive hurdles were also identified for the SIVcpz Vpu
637 protein (18, 37, 62). It has been shown that upon cross-species transmission, the ape
638 precursors of HIV-1 had to switch from Nef- to Vpu-mediated tetherin antagonism (37, 62).
639 However, only the pandemic M group viruses acquired efficient anti-tetherin activity, while the
640 much less prevalent group N, O and P viruses either failed to gain this activity or lost other Vpu
641 functions (37, 62, 63, 82, 86, 87). These findings have been taken to indicate that successful
642 SIV zoonoses require effective tetherin antagonism (23). When we compared ape virus Vpu
643 sequences to that of the HIV-1 group M consensus, we found that some SIVcpz*Ptt* Vpus
644 required only very few changes to gain key functional motifs (Fig. 8). In contrast, most
645 SIVcpz*Pts* Vpus required a substantially larger number of substitutions to acquire these same
646 human specific signatures (Fig. 8). Although counting numbers of mutational steps represents a
647 gross oversimplification of the adaptation process, our analyses suggest that certain SIVcpz*Ptt*
648 strains are better equipped to become new human pathogens than most SIVcpz*Pts* strains. The
649 fact that the reconstructed BF1167 genome (as well as other SIVcpz*Pts* strains) replicates well
650 in human CD4+ T cells (Fig. 5) does not argue against this, since such maximally stimulated
651 cultures do not accurately recapitulate the conditions of viral replication and transmission *in vivo*
652 (7).

653 In summary, we report here that wild-living *P. t. schweinfurthii* populations are much
654 more widely and commonly infected with SIVcpz than previously appreciated. This is particularly
655 true for communities in the northern DRC, which represent a large continuous population that
656 seems to provide opportunity for virus flow across vast areas. Whether these viruses have a

reduced potential to infect humans and cause epidemic outbreaks is not known, but should be investigated. In particular, studies of host restriction mechanisms that may have prevented the spread of these viruses in the human population would be informative. This is now possible, since well-characterized reagents, including a large panel of SIVcpz*Ptt* and SIVcpz*Pts* infectious molecular clones, are available for study. The high prevalence of SIVcpz*Pts* infection also has implications for conservation efforts. SIVcpz is quite pathogenic and has been shown to negatively impact chimpanzee population growth (17, 29, 55, 75). Given that the DRC is home to half of the world's remaining chimpanzees, it will be critical to determine in much greater detail to what extent SIVcpz has penetrated these populations, such that the impact of this infection on the long-term survival of this species can be determined. Finer grained prevalence and natural history data will also be critical for attempts to limit the spread of SIVcpz in wild ape populations, for example through the use of adeno-associated virus (AAV) mediated gene transfer of antibodies that neutralize SIVcpz as a prophylactic or therapeutic vaccine (4, 27), which could potentially be administered in wild settings (58).

## ACKNOWLEDGEMENTS

We thank Claude Kitoko and Bola Iyoka for collection of chimpanzee and bonobo fecal samples in the Parisi Forest and the Ikela region (DRC), respectively; the staff of the World Wildlife Fund (WWF) for collecting bonobo fecal samples in the Lac Tumba area (DRC); James Kakura, Geresomu Muhumuza, Monday Gideon and Raymond Ogen for field work in the Budongo Forest (Uganda); Sylvain Nyandwi, Thomas Safari, Samuel Uwimana, Patience Mwiseneza, Alex Ndayambaje, Isaac Ngayincyuro, Olivier Ngabonziza, and Eric Munyeshuli for field work in the Gishwati Reserve (Rwanda); Cleve Hicks for unpublished behavioral data from eastern chimpanzees in the northern DRC; Frank Kirchhoff and Daniel Sauter for helpful discussions; Patricia Crystal for artwork and preparation of the manuscript; the Ministry of

Scientific Research and Technology, the Department of Ecology and Management of Plant and Animal Resources of the University of Kisangani, the Ministries of Health and Environment and the National Ethics committee for permission to collect samples in the DRC; the Uganda Wildlife Authority and the Uganda National Council for Science and Technology for permission to conduct research in the Budongo Forest and the Kyambura Gorge of Queen Elizabeth National Park; the Rwandan Office of Tourism and National Parks for permission to collect samples in the Gishwati Forest Reserve. This work was supported in part by grants from the National Institutes of Health (R01 AI50529, R01 AI58715; P30 AI 27767), the Agence Nationale de Recherches sur le SIDA (ANRS 12255), and the Great Ape Trust; RSR was funded by a Howard Hughes Medical Institute Med-into-Grad Fellowship, and SAM by a grant from Infectiopole Sud, France.

## REFERENCES

1. **Abascal, F., R. Zardoya, and D. Posada.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:**2104-2105.
2. **Ahuka-Mundeke, S., A. Ayouba, P. Mbala-Kingebeni, F. Liegeois, A. Esteban, O. Lunguya-Metila, D. Demba, G. Bilulu, V. Mbenzo-Abokome, B. I. Inogwabini, J. J. Muyembe-Tamfum, E. Delaporte, and M. Peeters.** 2011. Novel multiplexed HIV/simian immunodeficiency virus antibody detection assay. Emerg Infect Dis **17:**2277-2286.
3. **Bailes, E., F. Gao, F. Bibollet-Ruche, V. Courgnaud, M. Peeters, P. A. Marx, B. H. Hahn, and P. M. Sharp.** 2003. Hybrid origin of SIV in chimpanzees. Science **300:**1713.
4. **Balazs, A. B., J. Chen, C. M. Hong, D. S. Rao, L. Yang, and D. Baltimore.** 2012. Antibody-based protection against HIV infection by vectored immunoprophylaxis. Nature **481:**81-84.

5. **Bennett, E. L., E. Blencowe, K. Brandon, D. Brown, R. W. Burn, G. Cowlishaw, G. Davies, H. Dublin, J. E. Fa, E. J. Milner-Gulland, J. G. Robinson, J. M. Rowcliffe, F. M. Underwood, and D. S. Wilkie.** 2007. Hunting for consensus: reconciling bushmeat harvest, conservation, and development policy in West and Central Africa. Conserv Biol **21:**884-887.

6. **Betsem, E., R. Rua, P. Tortevoye, A. Froment, and A. Gessain.** 2011. Frequent and recent human acquisition of simian foamy viruses through apes' bites in central Africa. PLoS Pathog **7:**e1002306.

7. **Bibollet-Ruche, F., A. Heigele, B. F. Keele, J. L. Easlick, J. M. Decker, J. Takehisa, G. Learn, P. M. Sharp, B. H. Hahn, and F. Kirchhoff.** 2012. Efficient SIVcpz replication in human lymphoid tissue requires viral matrix protein adaptation. J Clin Invest **122:**1644-1652.

8. **Bonifacino, J. S., and L. M. Traub.** 2003. Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem **72:**395-447.

9. **Bowden, R., T. S. MacFie, S. Myers, G. Hellenthal, E. Nerrienet, R. E. Bontrop, C. Freeman, P. Donnelly, and N. I. Mundy.** 2012. Genomic tools for evolution and conservation in the chimpanzee: *Pan troglodytes ellioti* is a genetically distinct population. PLoS Genet **8:**e1002504.

10. **Butynski, T. M.** 2001. Africa's great apes., p. 3-56. *In* B. Beck, T. Stoinski, M. Hutchins, T. Maple, B. Norton, A. Rowan, S. E., and A. Arluke (ed.), Great apes and humans - the ethics of coexistence. Smithsonian Institution Press, Washington, D.C.

11. **Calvignac-Spencer, S., S. A. J. Leendertz, T. R. Gillespie, and F. H. Leendertz.** 2012. Wild great apes as sentinels and sources of infectious disease. Clin Microbiol Infect **18:**521-527.

12. **Chancellor, R. L., K. Langergraber, S. Ramirez, A. S. Rundus, and L. Vigilant.** 2012. Genetic sampling of unhabituated chimpanzees (*Pan troglodytes schweinfurthii*) in

734 Gishwati forest reserve, an isolated forest fragment in western Rwanda. Int J Primatol
735 **33:**479-488.
736 13. **Chen, B. K., R. T. Gandhi, and D. Baltimore.** 1996. CD4 down-modulation during
737 infection of human T cells with human immunodeficiency virus type 1 involves
738 independent activities of *vpu*, *env*, and *nef*. J Virol **70:**6044-6053.
739 14. **Decker, J. M., K. P. Zammit, J. L. Easlick, M. L. Santiago, D. Bonenberger, B. H.**
740 **Hahn, O. Kutsch, and F. Bibollet-Ruche.** 2009. Effective activation alleviates the
741 replication block of CCR5-tropic HIV-1 in chimpanzee CD4+ lymphocytes. Virology
742 **394:**109-118.
743 15. **Dube, M., B. B. Roy, P. Guiot-Guillain, J. Binette, J. Mercier, A. Chiasson, and E. A.**
744 **Cohen.** 2010. Antagonism of tetherin restriction of HIV-1 release by Vpu involves
745 binding and sequestration of the restriction factor in a perinuclear compartment. PLoS
746 Pathog **6:**e1000856.
747 16. **Eng, B., P. Ainsworth, and J. S. Waye.** 1994. Anomalous migration of PCR products
748 using nondenaturing polyacrylamide gel electrophoresis: the amelogenin sex-typing
749 system. J Forensic Sci **39:**1356-1359.
750 17. **Etienne, L., E. Nerrienet, M. LeBreton, G. T. Bibila, Y. Foupouapouognigni, D.**
751 **Rousset, A. Nana, C. F. Djoko, U. Tamoufe, A. F. Aghokeng, E. Mpoudi-Ngole, E.**
752 **Delaporte, M. Peeters, N. D. Wolfe, and A. Ayouba.** 2011. Characterization of a new
753 simian immunodeficiency virus strain in a naturally infected *Pan troglodytes troglodytes*
754 chimpanzee with AIDS related symptoms. Retrovirology **8:**4.
755 18. **Evans, D. T., R. Serra-Moreno, R. K. Singh, and J. C. Guatelli.** 2010. BST-2/tetherin:
756 a new component of the innate immune response to enveloped viruses. Trends
757 Microbiol **18:**388-396.
758 19. **Fruth, B., J. M. Benishay, I. Bila-Isia, S. Coxe, J. Dupain, T. Furuichi, J. Hart, T. Hart,**
759 **C. Hashimoto, G. Hohmann, M. Hurley, O. Ilambu, M. Mulavwa, M. Ndunda, V.**

**Omasombo, G. Reinartz, J. Scherlis, L. Steel, and J. Thompson.** 2008, posting date. *Pan paniscus. In*: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. http://www.iucnredlist.org.

20. **Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature **397:**436-441.

21. **Gough, N. R., M. E. Zweifel, O. Martinez-Augustin, R. C. Aguilar, J. S. Bonifacino, and D. M. Fambrough.** 1999. Utilization of the indirect lysosome targeting pathway by lysosome-associated membrane proteins (LAMPs) is influenced largely by the C-terminal residue of their GYXXΦ targeting signals. J Cell Sci **112 (Pt 23):**4257-4269.

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

23. **Gupta, R. K., and G. J. Towers.** 2009. A tail of tetherin: how pandemic HIV-1 conquered the world. Cell Host Microbe **6:**393-395.

24. **Hahn, B. H., G. M. Shaw, K. M. De Cock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. Science **287:**607-614.

25. **Hicks, T. C.** 2010. A chimpanzee Mega-Culture? Exploring behavioral continuity in *Pan troglodytes schweinfurthii* across northern DR Congo. Ph.D. thesis. University of Amsterdam.

26. **Hicks, T. C., L. Darby, J. Hart, J. Swinkels, N. January, and S. Menken.** 2010. Trade in orphans and bushmeat threatens one of the Democratic Republic of the Congo's most important populations of eastern chimpanzees (*Pan troglodytes schweinfurthii*). Afr Primates **7:**1-18.

27. **Johnson, P. R., B. C. Schnepp, J. Zhang, M. J. Connell, S. M. Greene, E. Yuste, R. C. Desrosiers, and K. R. Clark.** 2009. Vector-mediated gene transfer engenders long-

lived neutralizing activity and protection against SIV infection in monkeys. Nat Med **15:**901-906.

28. **Kanemori, Y., K. Uto, and N. Sagata.** 2005. Beta-TrCP recognizes a previously undescribed nonphosphorylated destruction motif in Cdc25A and Cdc25B phosphatases. Proc Natl Acad Sci U S A **102:**6279-6284.

29. **Keele, B. F., J. H. Jones, K. A. Terio, J. D. Estes, R. S. Rudicell, M. L. Wilson, Y. Li, G. H. Learn, T. M. Beasley, J. Schumacher-Stankey, E. Wroblewski, A. Mosser, J. Raphael, S. Kamenya, E. V. Lonsdorf, D. A. Travis, T. Mlengeya, M. J. Kinsel, J. G. Else, G. Silvestri, J. Goodall, P. M. Sharp, G. M. Shaw, A. E. Pusey, and B. H. Hahn.** 2009. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature **460:**515-519.

30. **Keele, B. F., F. Van Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. H. Hahn.** 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science **313:**523-526.

31. **Kobayashi, T., H. Ode, T. Yoshida, K. Sato, P. Gee, S. P. Yamamoto, H. Ebina, K. Strebel, H. Sato, and Y. Koyanagi.** 2011. Identification of amino acids in the human tetherin transmembrane domain responsible for HIV-1 Vpu interaction and susceptibility. J Virol **85:**932-945.

32. **Krüger, O., E. Affeldt, M. Brackmann, and K. Milhahn.** 1998. Group size and composition of *Colobus guereza* in Kyambura Gorge, southwest Uganda, in relation to chimpanzee activity. Int J Primatol **19:**287-297.

33. **Kueck, T., and S. J. Neil.** 2012. A cytoplasmic tail determinant in HIV-1 Vpu mediates targeting of tetherin for endosomal degradation and counteracts interferon-induced restriction. PLoS Pathog **8:**e1002609.

34. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

35. **Leendertz, S. A., S. Locatelli, C. Boesch, C. Kucherer, P. Formenty, F. Liegeois, A. Ayouba, M. Peeters, and F. H. Leendertz.** 2011. No evidence for transmission of SIVwrc from western red colobus monkeys (*Piliocolobus badius badius*) to wild West African chimpanzees (*Pan troglodytes verus*) despite high exposure through hunting. BMC Microbiol **11:**24.

36. **Liegeois, F., V. Boué, S. Locatelli, C. Butel, A. Mouinga-Ondeme, E. Delaporte, J.-P. Gonzalez, M. Peeters, and F. Rouet.** 2012. Identification of new divergent SIVcpz*Ptt* strains in wild living chimpanzees in Gabon, The Conference on Retroviruses and Opportunistic Infections, Seattle, Washington.

37. **Lim, E. S., H. S. Malik, and M. Emerman.** 2010. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. J Virol **84:**7124-7134.

38. **Ling, B., M. L. Santiago, S. Meleth, B. Gormus, H. M. McClure, C. Apetrei, B. H. Hahn, and P. A. Marx.** 2003. Noninvasive detection of new simian immunodeficiency virus lineages in captive sooty mangabeys: ability to amplify virion RNA from fecal samples correlates with viral load in plasma. J Virol **77:**2214-2226.

39. **Liu, W., M. Worobey, Y. Li, B. F. Keele, F. Bibollet-Ruche, Y. Guo, P. A. Goepfert, M. L. Santiago, J. B. Ndjango, C. Neel, S. L. Clifford, C. Sanz, S. Kamenya, M. L. Wilson, A. E. Pusey, N. Gross-Camp, C. Boesch, V. Smith, K. Zamma, M. A. Huffman, J. C. Mitani, D. P. Watts, M. Peeters, G. M. Shaw, W. M. Switzer, P. M. Sharp, and B. H. Hahn.** 2008. Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. PLoS Pathog **4:**e1000097.

838 40. **Margottin, F., S. P. Bour, H. Durand, L. Selig, S. Benichou, V. Richard, D. Thomas,**
839 **K. Strebel, and R. Benarous.** 1998. A novel human WD protein, h-βTrCp, that interacts
840 with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif.
841 Mol Cell **1:**565-574.
842 41. **Moll, M., S. K. Andersson, A. Smed-Sorensen, and J. K. Sandberg.** 2010. Inhibition
843 of lipid antigen presentation in dendritic cells by HIV-1 Vpu interference with CD1d
844 recycling from endosomal compartments. Blood **116:**1876-1884.
845 42. **Neel, C., L. Etienne, Y. Li, J. Takehisa, R. S. Rudicell, I. N. Bass, J. Moudindo, A.**
846 **Mebenga, A. Esteban, F. Van Heuverswyn, F. Liegeois, P. J. Kranzusch, P. D.**
847 **Walsh, C. M. Sanz, D. B. Morgan, J. B. Ndjango, J. C. Plantier, S. Locatelli, M. K.**
848 **Gonder, F. H. Leendertz, C. Boesch, A. Todd, E. Delaporte, E. Mpoudi-Ngole, B. H.**
849 **Hahn, and M. Peeters.** 2010. Molecular epidemiology of simian immunodeficiency virus
850 infection in wild-living gorillas. J Virol **84:**1464-1476.
851 43. **Neil, S. J., T. Zang, and P. D. Bieniasz.** 2008. Tetherin inhibits retrovirus release and is
852 antagonized by HIV-1 Vpu. Nature **451:**425-430.
853 44. **Peeters, M., V. Courgnaud, B. Abela, P. Auzel, X. Pourrut, F. Bibollet-Ruche, S.**
854 **Loul, F. Liegeois, C. Butel, D. Koulagna, E. Mpoudi-Ngole, G. M. Shaw, B. H. Hahn,**
855 **and E. Delaporte.** 2002. Risk to human health from a plethora of simian
856 immunodeficiency viruses in primate bushmeat. Emerg Infect Dis **8:**451-457.
857 45. **Peeters, M., K. Fransen, E. Delaporte, M. Van den Haesevelde, G. M. Gershy-Damet,**
858 **L. Kestens, G. van der Groen, and P. Piot.** 1992. Isolation and characterization of a
859 new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-
860 captured chimpanzee. Aids **6:**447-451.
861 46. **Pepin, J.** 2011. The origins of AIDS. Cambridge University Press, New York.

47. **Perez-Caballero, D., T. Zang, A. Ebrahimi, M. W. McNatt, D. A. Gregory, M. C. Johnson, and P. D. Bieniasz.** 2009. Tetherin inhibits HIV-1 release by directly tethering virions to cells. Cell **139:**499-511.

48. **Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemee, F. Damond, D. L. Robertson, and F. Simon.** 2009. A new human immunodeficiency virus derived from gorillas. Nat Med **15:**871-872.

49. **Platt, E. J., K. Wehrly, S. E. Kuhmann, B. Chesebro, and D. Kabat.** 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol **72:**2855-2864.

50. **Plumptre, A. J., M. Masozera, and A. Vedder.** 2001. The impact of civil war on the conservation of protected areas in Rwanda. Washington, D.C.: Biodiversity Support Program.

51. **Plumptre, A. J., R. Rose, G. Nangendo, E. A. Williamson, K. Didier, J. Hart, F. Mulindahabi, C. Hicks, B. Griffin, H. Ogawa, S. Nixon, L. Pintea, A. Vosper, M. McClennan, F. Amsini, A. McNeilage, J. R. Makana, M. Kanamori, A. Hernandez, A. Piel, F. Stewart, J. Moore, K. Zamma, M. Nakamura, S. Kamenya, G. Idani, T. Sakamaki, M. Yoshikawa, D. Greer, S. Tranquilli, R. Beyers, T. Furuichi, C. Hashimoto, and E. Bennett.** 2010. Eastern chimpanzee (*Pan troglodytes schweinfurthii*): status survey and conservation action plan 2010-2020. IUCN/SSC Primate Specialist Group.

52. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817-818.

53. **Reynolds, V.** 2005. The chimpanzees of the Budongo forest: ecology, behaviour, and conservation. Oxford University Press, New York.

54. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-1574.

55. **Rudicell, R. S., J. Holland Jones, E. E. Wroblewski, G. H. Learn, Y. Li, J. D. Robertson, E. Greengrass, F. Grossmann, S. Kamenya, L. Pintea, D. C. Mjungu, E. V. Lonsdorf, A. Mosser, C. Lehman, D. A. Collins, B. F. Keele, J. Goodall, B. H. Hahn, A. E. Pusey, and M. L. Wilson.** 2010. Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. PLoS Pathog **6:**e1001116.

56. **Rudicell, R. S., A. K. Piel, F. Stewart, D. L. Moore, G. H. Learn, Y. Li, J. Takehisa, L. Pintea, G. M. Shaw, J. Moore, P. M. Sharp, and B. H. Hahn.** 2011. High prevalence of simian immunodeficiency virus infection in a community of savanna chimpanzees. J Virol **85:**9918-9928.

57. **Ruiz, A., M. S. Hill, K. Schmitt, J. Guatelli, and E. B. Stephens.** 2008. Requirements of the membrane proximal tyrosine and dileucine-based sorting signals for efficient transport of the subtype C Vpu protein to the plasma membrane and in virus release. Virology **378:**58-68.

58. **Ryan, S. J., and P. D. Walsh.** 2011. Consequences of non-intervention for infectious disease in African great apes. PLoS One **6:**e29030.

59. **Santiago, M. L., F. Bibollet-Ruche, E. Bailes, S. Kamenya, M. N. Muller, M. Lukasik, A. E. Pusey, D. A. Collins, R. W. Wrangham, J. Goodall, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 2003. Amplification of a complete simian immunodeficiency virus genome from fecal RNA of a wild chimpanzee. J Virol **77:**2233-2242.

60. **Santiago, M. L., M. Lukasik, S. Kamenya, Y. Li, F. Bibollet-Ruche, E. Bailes, M. N. Muller, M. Emery, D. A. Goldenberg, J. S. Lwanga, A. Ayouba, E. Nerrienet, H. M. McClure, J. L. Heeney, D. P. Watts, A. E. Pusey, D. A. Collins, R. W. Wrangham, J. Goodall, J. F. Brookfield, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2003. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). J Virol **77:**7545-7562.

61. **Santiago, M. L., C. M. Rodenburg, S. Kamenya, F. Bibollet-Ruche, F. Gao, E. Bailes, S. Meleth, S. J. Soong, J. M. Kilby, Z. Moldoveanu, B. Fahey, M. N. Muller, A. Ayouba, E. Nerrienet, H. M. McClure, J. L. Heeney, A. E. Pusey, D. A. Collins, C. Boesch, R. W. Wrangham, J. Goodall, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2002. SIVcpz in wild chimpanzees. Science **295:**465.

62. **Sauter, D., M. Schindler, A. Specht, W. N. Landford, J. Munch, K. A. Kim, J. Votteler, U. Schubert, F. Bibollet-Ruche, B. F. Keele, J. Takehisa, Y. Ogando, C. Ochsenbauer, J. C. Kappes, A. Ayouba, M. Peeters, G. H. Learn, G. Shaw, P. M. Sharp, P. Bieniasz, B. H. Hahn, T. Hatziioannou, and F. Kirchhoff.** 2009. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. Cell Host Microbe **6:**409-421.

63. **Sauter, D., A. Specht, and F. Kirchhoff.** 2010. Tetherin: holding on and letting go. Cell **141:**392-398.

64. **Shah, A. H., B. Sowrirajan, Z. B. Davis, J. P. Ward, E. M. Campbell, V. Planelles, and E. Barker.** 2010. Degranulation of natural killer cells following interaction with HIV-1-infected cells is hindered by downmodulation of NTB-A by Vpu. Cell Host Microbe **8:**397-409.

65. **Sharp, P. M., and B. H. Hahn.** 2010. The evolution of HIV-1 and the origin of AIDS. Philos T R Soc B **365:**2487-2494.

66. **Sharp, P. M., and B. H. Hahn.** 2011. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med **1:**a006841.

67. **Sharp, P. M., G. M. Shaw, and B. H. Hahn.** 2005. Simian immunodeficiency virus infection of chimpanzees. J Virol **79:**3891-3902.

68. **Simon, F., P. Mauclere, P. Roques, I. Loussert-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi, and F. Brun-Vezinet.** 1998.

Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med **4:**1032-1037.

69. **Skasko, M., Y. Wang, Y. Tian, A. Tokarev, J. Munguia, A. Ruiz, E. B. Stephens, S. J. Opella, and J. Guatelli.** 2012. HIV-1 Vpu protein antagonizes innate restriction factor BST-2 via lipid-embedded helix-helix interactions. J Biol Chem **287:**58-67.

70. **Sullivan, K. M., A. Mannucci, C. P. Kimpton, and P. Gill.** 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques **15:**636-638, 640-631.

71. **Surbeck, M., A. Fowler, C. Deimel, and G. Hohmann.** 2009. Evidence for the consumption of arboreal, diurnal primates by bonobos (*Pan paniscus*). Am J Primatol **71:**171-174.

72. **Surbeck, M., and G. Hohmann.** 2008. Primate hunting by bonobos at LuiKotale, Salonga National Park. Curr Biol **18:**R906-907.

73. **Takehisa, J., M. H. Kraus, A. Ayouba, E. Bailes, F. Van Heuverswyn, J. M. Decker, Y. Li, R. S. Rudicell, G. H. Learn, C. Neel, E. M. Ngole, G. M. Shaw, M. Peeters, P. M. Sharp, and B. H. Hahn.** 2009. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J Virol **83:**1635-1648.

74. **Takehisa, J., M. H. Kraus, J. M. Decker, Y. Li, B. F. Keele, F. Bibollet-Ruche, K. P. Zammit, Z. Weng, M. L. Santiago, S. Kamenya, M. L. Wilson, A. E. Pusey, E. Bailes, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2007. Generation of infectious molecular clones of simian immunodeficiency virus from fecal consensus sequences of wild chimpanzees. J Virol **81:**7463-7475.

75. **Terio, K. A., M. J. Kinsel, J. Raphael, T. Mlengeya, I. Lipende, C. A. Kirchhoff, B. Gilagiza, M. L. Wilson, S. Kamenya, J. D. Estes, B. F. Keele, R. S. Rudicell, W. Liu, S. Patton, A. Collins, B. H. Hahn, D. A. Travis, and E. V. Lonsdorf.** 2011. Pathologic

lesions in chimpanzees (*Pan trogylodytes schweinfurthii*) from Gombe National Park, Tanzania, 2004-2010. J Zoo Wildl Med **42:**597-607.

76. **Van Damme, N., D. Goff, C. Katsura, R. L. Jorgenson, R. Mitchell, M. C. Johnson, E. B. Stephens, and J. Guatelli.** 2008. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe **3:**245-252.

77. **Van Dooren, S., W. M. Switzer, W. Heneine, P. Goubau, E. Verschoor, B. Parekh, W. De Meurichy, C. Furley, M. Van Ranst, and A. M. Vandamme.** 2002. Lack of evidence for infection with simian immunodeficiency virus in bonobos. AIDS Res Hum Retroviruses **18:**213-216.

78. **Van Heuverswyn, F., Y. Li, E. Bailes, C. Neel, B. Lafay, B. F. Keele, K. S. Shaw, J. Takehisa, M. H. Kraus, S. Loul, C. Butel, F. Liegeois, B. Yangda, P. M. Sharp, E. Mpoudi-Ngole, E. Delaporte, B. H. Hahn, and M. Peeters.** 2007. Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. Virology **368:**155-171.

79. **Van Heuverswyn, F., Y. Li, C. Neel, E. Bailes, B. F. Keele, W. Liu, S. Loul, C. Butel, F. Liegeois, Y. Bienvenue, E. M. Ngolle, P. M. Sharp, G. M. Shaw, E. Delaporte, B. H. Hahn, and M. Peeters.** 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature **444:**164.

80. **Vanden Haesevelde, M. M., M. Peeters, G. Jannes, W. Janssens, G. van der Groen, P. M. Sharp, and E. Saman.** 1996. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. Virology **221:**346-350.

81. **Vigan, R., and S. J. Neil.** 2010. Determinants of tetherin antagonism in the transmembrane domain of the human immunodeficiency virus type 1 Vpu protein. J Virol **84:**12958-12970.

82. **Vigan, R., and S. J. Neil.** 2011. Separable determinants of subcellular localization and interaction account for the inability of group O HIV-1 Vpu to counteract tetherin. J Virol **85:**9737-9748.

83. **Wain, L. V., E. Bailes, F. Bibollet-Ruche, J. M. Decker, B. F. Keele, F. Van Heuverswyn, Y. Li, J. Takehisa, E. M. Ngole, G. M. Shaw, M. Peeters, B. H. Hahn, and P. M. Sharp.** 2007. Adaptation of HIV-1 to its human host. Mol Biol Evol **24:**1853-1860.

84. **Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw.** 2003. Antibody neutralization and escape by HIV-1. Nature **422:**307-312.

85. **Worobey, M., M. L. Santiago, B. F. Keele, J. B. Ndjango, J. B. Joy, B. L. Labama, A. B. Dhed, A. Rambaut, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2004. Origin of AIDS: contaminated polio vaccine theory refuted. Nature **428:**820.

86. **Yang, S. J., L. A. Lopez, C. M. Exline, K. G. Haworth, and P. M. Cannon.** 2011. Lack of adaptation to human tetherin in HIV-1 group O and P. Retrovirology **8:**78.

87. **Yang, S. J., L. A. Lopez, H. Hauser, C. M. Exline, K. G. Haworth, and P. M. Cannon.** 2010. Anti-tetherin activities in Vpu-expressing primate lentiviruses. Retrovirology **7:**13.

## FIGURE LEGENDS

**FIG 1** Location of ape study sites. (A) Geographic ranges of chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) in sub-Saharan Africa. The four recognized chimpanzee subspecies are color-coded (*P. t. verus*, grey; *P. t. ellioti*; magenta; *P. t. troglodytes*; blue; *P. t. schweinfurthii*, yellow). International borders, major rivers and lakes, and select cities are shown. Asterisks

indicate where the closest SIVcpz relatives of HIV-1 groups M (red) and N (green) were identified in wild-living *P. t. troglodytes* communities (30). A box outlines the study area, which is magnified in panel B. (B) Location of chimpanzee (circles) and bonobo (squares) study sites in the DRC, Uganda and Rwanda. The ranges of eastern chimpanzees (yellow) and bonobos (orange) are shown as in (A). Sites where SIVcpz was detected are indicated in red, with white and yellow lettering denoting the recovery of antibody positive versus antibody and nucleic acid positive samples, respectively. Previously published SIVcpz positive and negative sites in Uganda, Rwanda and Tanzania are shown in dark red and gray, respectively (GM, Gombe National Park; UG, Masito-Ugalla region). Forested areas are shown in green, while arid and semi-arid areas are in yellow and brown. Major lakes are shown in black with major rivers depicted in blue. Dashed white lines indicate national boundaries.

**FIG 2** Detection of SIVcpz antibodies in chimpanzee fecal samples. Fecal samples from eastern chimpanzees (middle) and bonobos (right) as well as human controls (left) were tested by enhanced chemiluminescent Western blot using HIV-1 antigen containing strips. Samples are numbered, with letters indicating their collection site as shown in Fig. 1B. Molecular weights of HIV-1 proteins are indicated. The banding pattern of plasma from HIV-1 infected (positive) and uninfected (negative) humans are shown for control.

**FIG 3** SIVcpz strains from the DRC cluster according to their subspecies of origin. A maximum likelihood tree was constructed from partial (232 bp) *pol* sequences (spanning HXB2 coordinates 4682 – 4913). Newly characterized SIVcpz strains from the DRC are highlighted, with sequences from the same individual color-coded (for individual designation and sample numbers see Table S3). Previously characterized SIVcpz, SIVgor and HIV-1 strains forming the SIVcpz*Ptt* (top cluster) and SIVcpz*Pts* (bottom cluster) lineages are shown in black. The latter include reference strains from Gombe (TAN1, TAN2, TAN3, TAN5, TAN13) and Ugalla (UG38),

as well as ANT, which is of unknown origin. Asterisks indicate bootstrap support ≥ 70%. The scale bar represents 0.05 substitutions per site.

**FIG 4** Phylogeny of SIVcpz in the DRC. Maximum likelihood trees were constructed of partial (A) *pol* (HXB2 coordinates 3887 – 4778, (B) *vpu*/*env* (HXB2 coordinates 6062 – 6578), (C) gp41 (HXB2 coordinates 7836 – 8264, and gp41/*nef* (HXB2 coordinates 8277 – 9047) sequences. Regions of ambiguous alignment were removed from this analysis. New SIVcpz*Pts* strains from the DRC are show in blue, followed by the sample code in parentheses. Previously characterized SIVcpz, SIVgor and HIV-1 strains forming the SIVcpz*Ptt* (top cluster) and SIVcpz*Pts* (bottom cluster) lineages are shown in black. Nodes with both bootstrap support ≥ 70% and Bayesian posterior probability ≥ 0.95 are indicated by asterisks. Scale bar represents 0.05 substitutions/site.

**FIG 5** Generation and biological characterization of a replication competent SIVcpz*Pts* molecular clone. (A) Individual RT-PCR amplicons (orange boxes) of BF1167 are shown in relation to the SIVcpz genome. Fragments are drawn to scale, with nucleotide sequences numbered starting at the beginning of the R region in the 5' LTR (see scale bar). Three subgenomic fragments bound by *Mlu*I, *Nco*I, *Sal*I, and *Apa*I restriction sites were synthesized and then assembled to produce a full-length provirus (blue line). (B) The replication kinetics of BF1167 derived virus in human (top) and chimpanzee (bottom) CD4+ T-cells is shown in relation to HIV-1 (SG3, blue) and SIVcpz*Pts* (TAN2, green) reference strains (x-axis: days post infection; y-axis nanogram of reverse transcriptase (RT) activity per ml of culture supernatant). Average values (and one standard deviation) from different experiments (indicated in parentheses) are shown. (C) TZM-bl cells were pretreated with AMD3100 (inhibitor of CXCR4), TAK779 (inhibitor of CCR5), or both prior to addition of the virus preparations indicated. Virus infectivity is plotted on the vertical axis as a percentage of the untreated control. Virus derived

from reference clones NL4.3 (X4-tropic), YU2 (R5-tropic), and WEAU1.6 (dual tropic) as well as SIVcpz*Ptt* (MB897) and SIVcpz*Pts* (TAN2) strains were included for control. BF1167 is an R5-tropic virus.

**FIG 6** Evolutionary relationships of BF1167 full-length genome sequences. Maximum likelihood trees were inferred from amino acid (AA) sequence alignments of the major proteins, including (A) Gag (420 AA; HXB2 coordinates 790 – 2280), (B) N-terminal Pol (630 AA; HXB2 coordinates 2295 – 4184), (C) C-terminal Pol/Vif (453 AA; HXB2 coordinates 4185 – 5556), and (D) Env (729 AA; HXB2 coordinates 6324 – 8792); the Pol protein was separated into two fragments at a point where a recombination breakpoint was previously identified in HIV-1 group N. The BF1167 sequence is shown in blue. Previously characterized SIVcpz, SIVgor and HIV-1 strains forming the SIVcpz*Ptt* (top cluster) and SIVcpz*Pts* (bottom cluster) lineages are shown in black. Nodes with both bootstrap support ≥ 70% and Bayesian posterior probability ≥ 0.95 are indicated by asterisks. Scale bar represents 0.05 amino acid replacements/site.

**FIG 7** Adaptive requirements of SIVcpz at Gag-30. The codon at position 30 of the Gag matrix protein is shown for 14 SIVcpz*Pts* strains, including seven new viruses from the DRC. Ten of these 14 viruses encoded a Leu using TTA or TTG codons at Gag-30, which require at least two nucleotide changes to become Arg (CGN or AGR) or Lys (AAR) codons. In contrast, all sequenced SIVcpz*Ptt* (n=15) and SIVgor (n=3) strains contain a Met ATG codon, which requires only a single substitution to change to either human specific signature.

**FIG 8** Adaptive requirements of SIVcpz in transmembrane and cytoplasmic domains of Vpu. Vpu amino acid sequences of SIVcpz*Ptt*, SIVgor and SIVcpz*Pts* strains are aligned in their transmembrane domain to the corresponding region of the HIV-1 group M consensus as previously described (37). Dashes indicate gaps introduced to optimize the alignment. Gray

boxes highlight residues of a conserved helix-helix interaction motif (G/AxxxAxxxAxxxW; where 'x' may be any amino acid) that is required to counteract human tetherin (69, 81). The minimum number of mutational steps needed to change the corresponding amino acid to that of the human residue is indicated on the right. Columns on the far right indicate the presence (+) or absence (blank) of previously described transport and/or degradation motifs in the intracellular domain of Vpu. These include a YxxΦ motif scored as Yxx(L/M/V/I/F/W) (8, 21, 57), a β-TrCP ubiquitin-dependent degradation signal scored as D(S/D/E)Gxx(S/D/E) (28, 40), and a putative trafficking signal scored as (D/E)xxxL(L/V/I/M) (33).

**Table 1. Prevalence of SIVcpz infection in wild-living eastern chimpanzees and bonobos**

| Field site[1] | Country[2] | Species/ subspecies[3] | Fecal samples collected | Fecal samples positive for mtDNA | Proportion degraded/ mixed samples | Number of individuals sampled[4] | Number of mtDNA haplotypes[5] | SIVcpz antibody positive samples | Number of infected individuals[6] | vRNA positive samples | SIVcpz prevalence (%)[7] | 95% confidence interval |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AM | DRC | *P.t.s.* | 44 | 37 | 0.16 | n/a | 5 | 15 | 3 | 3 | 33 | 2-43 |
| AN | DRC | *P.t.s.* | 16 | 15 | 0.06 | n/a | 1 | 0 | 0 | 0 | 0 | 0-60 |
| AZ | DRC | *P.t.s.* | 6 | 5 | 0.17 | 3 | 2 | 4 | 2 | 0 | 67 | 9-99 |
| BA | DRC | *P.t.s.* | 256 | 229 | 0.11 | n/a | 22 | 39 | 5 | 1 | 9 | 3-20 |
| BD | DRC | *P.t.s.* | 15 | 15 | 0.00 | n/a | 9 | 0 | 0 | 0 | 0 | 0-34 |
| BE | DRC | *P.t.s.* | 134 | 87 | 0.35 | n/a | 18 | 0 | 0 | 0 | 0 | 0-17 |
| BF | DRC | *P.t.s.* | 42 | 37 | 0.12 | n/a | 7 | 12 | 1 | 12 | 11 | 0-48 |
| BI | DRC | *P.t.s.* | 124 | 103 | 0.17 | n/a | 14 | 0 | 0 | 0 | 0 | 0-14 |
| BL | DRC | *P.t.s.* | 50 | 48 | 0.04 | n/a | 21 | 3 | 2 | 2 | 10 | 1-30 |
| BM | DRC | *P.t.s.* | 47 | 43 | 0.09 | n/a | 2 | 0 | 0 | 0 | 0 | 0-31 |
| BR | DRC | *P.t.s.* | 59 | 59 | 0.00 | n/a | 1 | 0 | 0 | 0 | 0 | 0-23 |
| BS | DRC | *P.t.s.* | 20 | 10 | 0.50 | n/a | 8 | 0 | 0 | 0 | 0 | 0-37 |
| BT | DRC | *P.t.s.* | 5 | 4 | 0.20 | n/a | 4 | 0 | 0 | 0 | 0 | 0-60 |
| BU[8] | DRC | *P.t.s.* | 1[8] | 1 | 0.00 | n/a | 1 | 1 | 1 | 0 | 100 | 3-100 |
| DL | DRC | *P.t.s.* | 5 | 5 | 0.00 | n/a | 1 | 0 | 0 | 0 | 0 | 0-98 |
| EP | DRC | *P.t.s.* | 160 | 126 | 0.21 | n/a | 19 | 7 | 2 | 4 | 7 | 1-22 |
| GO | DRC | *P.t.s.* | 5 | 2 | 0.60 | n/a | 2 | 0 | 0 | 0 | 0 | 0-84 |
| IJ | DRC | *P.t.s.* | 27 | 7 | 0.74 | n/a | 2 | 0 | 0 | 0 | 0 | 0-84 |
| IS | DRC | *P.t.s.* | 5 | 4 | 0.20 | n/a | 3 | 0 | 0 | 0 | 0 | 0-71 |
| KA | DRC | *P.t.s.* | 164 | 126 | 0.23 | n/a | 23 | 33 | 13 | 10 | 43 | 25-63 |
| KE | DRC | *P.t.s.* | 15 | 7 | 0.53 | n/a | 3 | 0 | 0 | 0 | 0 | 0-71 |
| KO | DRC | *P.t.s.* | 90 | 78 | 0.13 | n/a | 12 | 9 | 1 | 0 | 6 | 0-27 |
| KS[9] | DRC | *P.t.s.* | 13[9] | 11 | 0.15 | n/a | 8 | 4 | 3 | 0 | 38 | 9-76 |
| LU | DRC | *P.t.s.* | 212 | 163 | 0.23 | n/a | 18 | 34 | 3 | 27 | 8 | 2-21 |
| MA | DRC | *P.t.s.* | 11 | 7 | 0.36 | n/a | 1 | 0 | 0 | 0 | 0 | 0-84 |
| MN | DRC | *P.t.s.* | 2 | 1 | 0.50 | n/a | 1 | 0 | 0 | 0 | 0 | 0-98 |
| MO | DRC | *P.t.s.* | 21 | 15 | 0.29 | n/a | 10 | 2 | 1 | 2 | 10 | 0-45 |
| MU | DRC | *P.t.s.* | 58 | 53 | 0.09 | n/a | 11 | 26 | 6 | 3 | 50 | 21-79 |
| NI | DRC | *P.t.s.* | 7 | 4 | 0.43 | n/a | 2 | 0 | 0 | 0 | 0 | 0-84 |
| ON | DRC | *P.t.s.* | 59 | 40 | 0.32 | n/a | 4 | 0 | 0 | 0 | 0 | 0-34 |
| OP | DRC | *P.t.s.* | 85 | 61 | 0.28 | n/a | 9 | 26 | 2 | 18 | 14 | 2-43 |
| PA | DRC | *P.t.s.* | 131 | 110 | 0.16 | n/a | 16 | 48 | 12 | 5 | 46 | 27-67 |
| PO | DRC | *P.t.s.* | 11 | 10 | 0.09 | n/a | 6 | 0 | 0 | 0 | 0 | 0-46 |
| RU | DRC | *P.t.s.* | 18 | 13 | 0.28 | n/a | 4 | 0 | 0 | 0 | 0 | 0-60 |
| UB | DRC | *P.t.s.* | 117 | 99 | 0.15 | n/a | 23 | 11 | 7 | 3 | 30 | 13-53 |
| UD | DRC | *P.t.s.* | 5 | 5 | 0.00 | n/a | 2 | 0 | 0 | 0 | 0 | 0-84 |
| UM | DRC | *P.t.s.* | 100 | 61 | 0.39 | n/a | 10 | 0 | 0 | 0 | 0 | 0-23 |
| WA | DRC | *P.t.s.* | 170 | 151 | 0.11 | n/a | 25 | 40 | 8 | 1 | 22 | 10-39 |
| WB | DRC | *P.t.s.* | 91 | 69 | 0.24 | n/a | 11 | 8 | 3 | 1 | 19 | 4-46 |
| WK | DRC | *P.t.s.* | 35 | 33 | 0.06 | n/a | 10 | 0 | 0 | 0 | 0 | 0-31 |
| WL | DRC | *P.t.s.* | 44 | 39 | 0.11 | n/a | 13 | 1 | 1 | 0 | 8 | 0-36 |
| BG | Uganda | *P.t.s.* | 20 | 19 | 0.05 | 19 | 9 | 0 | 0 | 0 | 0 | n/a |
| KY | Uganda | *P.t.s.* | 16 | 16 | 0.00 | 13 | 4 | 0 | 0 | 0 | 0 | n/a |
| GI | Rwanda | *P.t.s.* | 49 | 42 | 0.14 | n/a | 6 | 0 | 0 | 0 | 0 | 0-31 |
| **n=44** | | | **2,565** | **2,070** | **0.19** | **567**[10] | **383** | **323** | **76** | **92** | **13.4** | **10.7-16.5** |
| BJ | DRC | *P.p.* | 2 | 2 | 0 | n/a | 1 | 0 | 0 | 0 | 0 | 0-98 |
| BN | DRC | *P.p.* | 96 | 85 | 0.12 | n/a | 7 | 0 | 0 | 0 | 0 | 0-8 |
| IK | DRC | *P.p.* | 56 | 39 | 0.30 | 17 | 7 | 0 | 0 | 0 | 0 | 0-20 |
| KR | DRC | *P.p.* | 78 | 69 | 0.12 | 38 | 12 | 0 | 0 | 0 | 0 | 0-9 |
| LK | DRC | *P.p.* | 43 | 38 | 0.12 | 17 | 8 | 0 | 0 | 0 | 0 | 0-20 |
| ML | DRC | *P.p.* | 268 | 262 | 0.02 | n/a | 6 | 0 | 0 | 0 | 0 | 0-3 |
| **n=6** | | | **543** | **495** | **0.09** | **244**[10] | **41** | **0** | **0** | **0** | **0** | **0-1.5** |

[1]Field sites are designated by a two-letter code (their location is shown in Figure 1).
[2]DRC, Democratic Republic of the Congo.
[3]*P.t.s.*, *Pan troglodytes schweinfurthii*; *P.p.*, *Pan paniscus*.
[4]Chimpanzees at the BG and KY field sites were individually known; For the AZ, IK, KR and LK field sites, the number of sampled apes was determined by microsatellite analyses (Table S4); n/a; not available.
[5]Minimum number of individuals as indicated by the number of distinct mtDNA haplotypes.
[6]SIVcpz infected chimpanzees were enumerated for each field site by microsatellite analysis of antibody positive fecal samples (Table S3).
[7]Prevalence of SIVcpz infection (%) with 95% confidence intervals; values are based on the proportion of SIVcpz antibody positive fecal samples, corrected for sample degradation and oversampling.
[8]The single sample from the BU field site was collected from a pet chimpanzee.
[9]Samples from the KS field site include three from pet chimpanzees.
[10]Estimated number of sampled individuals (see methods for details).

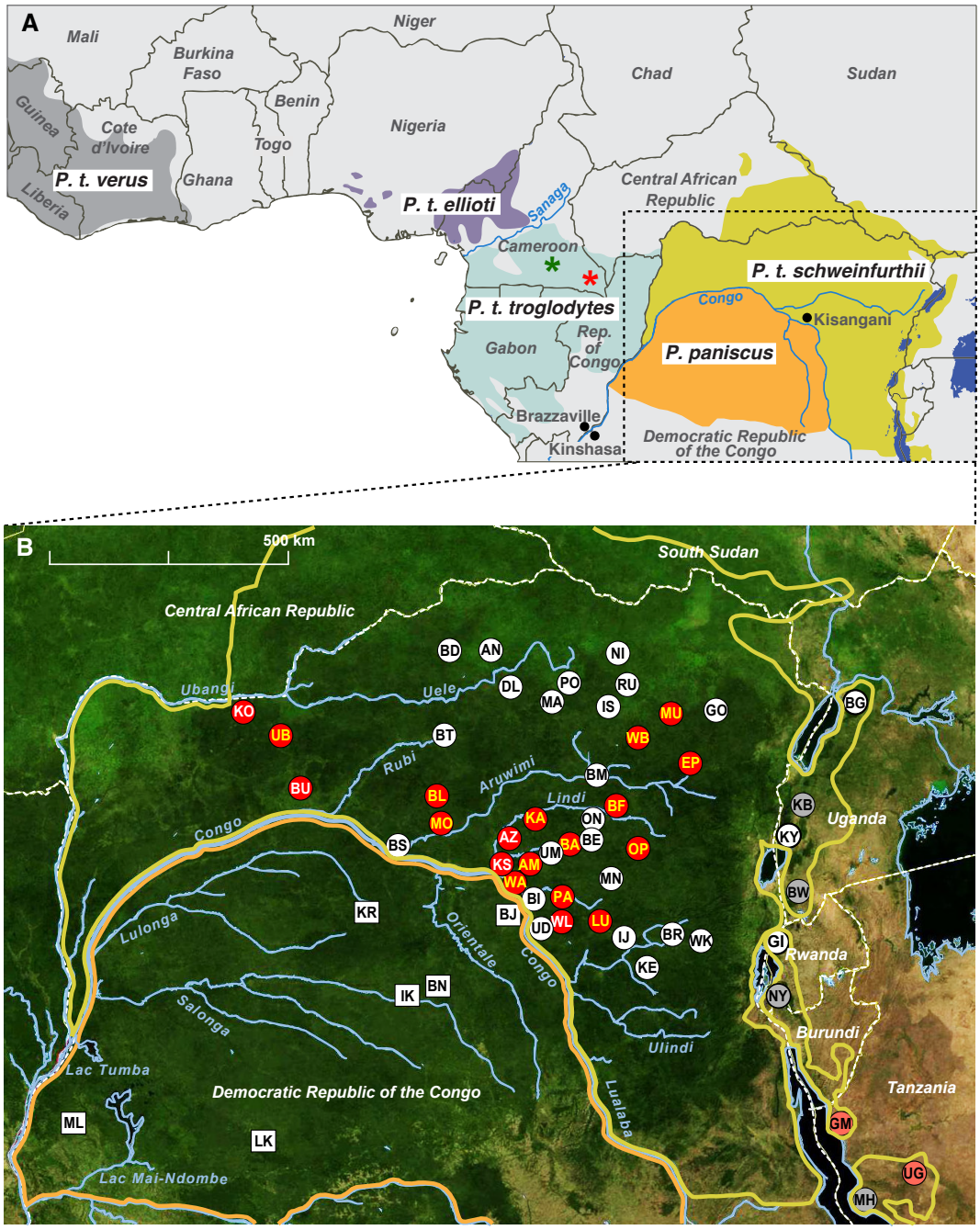
A
Mali
Burkina Faso
Niger
Chad
Sudan
Guinea
Cote d'Ivoire
Benin
Togo
Nigeria
Liberia
Ghana
P. t. verus
P. t. ellioti
Sanaga
Central African Republic
Cameroon
P. t. schweinfurthii
Congo
Kisangani
P. t. troglodytes
Rep. of Congo
Gabon
P. paniscus
Brazzaville
Kinshasa
Democratic Republic of the Congo
B
500 km
South Sudan
Central African Republic
Ubangi
Uele
Rubi
Aruwimi
Lindi
Congo
Lulonga
Salonga
Lac Tumba
Lac Mai-Ndombe
Orientale
Ulindi
Lualaba
Uganda
Rwanda
Burundi
Tanzania
Democratic Republic of the Congo
BD
AN
NI
DL
PO
RU
MA
IS
MU
GO
BG
KO
UB
BT
WB
EP
BU
BM
BL
BF
KA
ON
MO
AZ
BA
BE
OP
BS
KS
AM
JM
WA
MN
BI
PA
KR
BJ
UD
WL
LU
IJ
BR
WK
KE
IK
BN
ML
LK
KB
KY
BW
GI
NY
GM
UG
MH

**Figure 1**

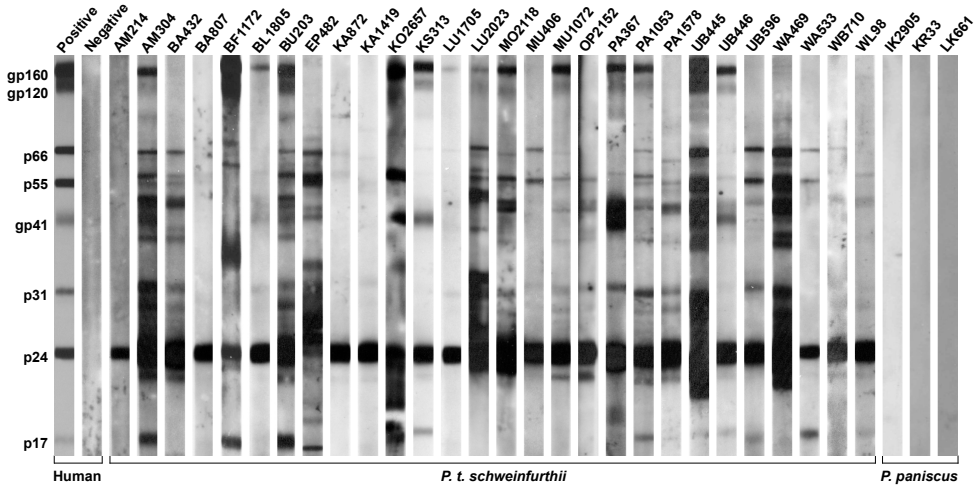
Positive
Negative
AM214
AM304
BA432
BA807
BF1172
BL1605
BU203
EP482
KA872
KA1419
KO2657
KS313
LU1705
LU2023
MO2118
MU406
MU1072
OP2152
PA367
PA1053
PA1578
UB445
UB446
UB596
WA469
WA533
WB710
WL98
IK2905
KR33
LK661
gp160
gp120
p66
p55
gp41
p31
p24
p17
Human
P. t. schweinfurthii
P. paniscus

**Figure 2**

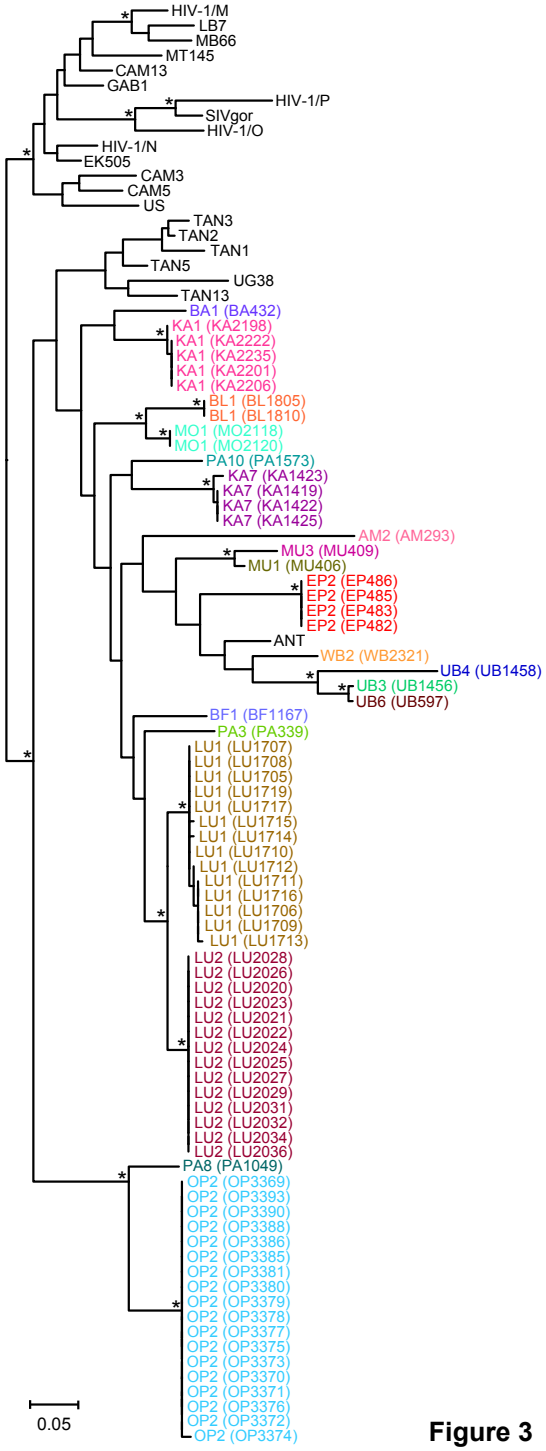
HIV-1/M
LB7
MB66
MT145
CAM13
GAB1
HIV-1/P
SIVgor
HIV-1/O
HIV-1/N
EK505
CAM3
CAM5
US
TAN3
TAN2
TAN1
TAN5
UG38
TAN13
BA1 (BA432)
KA1 (KA2198)
KA1 (KA2222)
KA1 (KA2235)
KA1 (KA2201)
KA1 (KA2206)
BL1 (BL1805)
BL1 (BL1810)
MO1 (MO2118)
MO1 (MO2120)
PA10 (PA1573)
KA7 (KA1423)
KA7 (KA1419)
KA7 (KA1422)
KA7 (KA1425)
AM2 (AM293)
MU3 (MU409)
MU1 (MU406)
EP2 (EP486)
EP2 (EP485)
EP2 (EP483)
EP2 (EP482)
ANT
WB2 (WB2321)
UB4 (UB1458)
UB3 (UB1456)
UB6 (UB597)
BF1 (BF1167)
PA3 (PA339)
LU1 (LU1707)
LU1 (LU1708)
LU1 (LU1705)
LU1 (LU1719)
LU1 (LU1717)
LU1 (LU1715)
LU1 (LU1714)
LU1 (LU1710)
LU1 (LU1712)
LU1 (LU1711)
LU1 (LU1716)
LU1 (LU1706)
LU1 (LU1709)
LU1 (LU1713)
LU2 (LU2028)
LU2 (LU2026)
LU2 (LU2020)
LU2 (LU2023)
LU2 (LU2021)
LU2 (LU2022)
LU2 (LU2024)
LU2 (LU2025)
LU2 (LU2027)
LU2 (LU2029)
LU2 (LU2031)
LU2 (LU2032)
LU2 (LU2034)
LU2 (LU2036)
PA8 (PA1049)
OP2 (OP3369)
OP2 (OP3393)
OP2 (OP3390)
OP2 (OP3388)
OP2 (OP3386)
OP2 (OP3385)
OP2 (OP3381)
OP2 (OP3380)
OP2 (OP3379)
OP2 (OP3378)
OP2 (OP3377)
OP2 (OP3375)
OP2 (OP3373)
OP2 (OP3370)
OP2 (OP3371)
OP2 (OP3376)
OP2 (OP3372)
OP2 (OP3374)
0.05

**Figure 3**

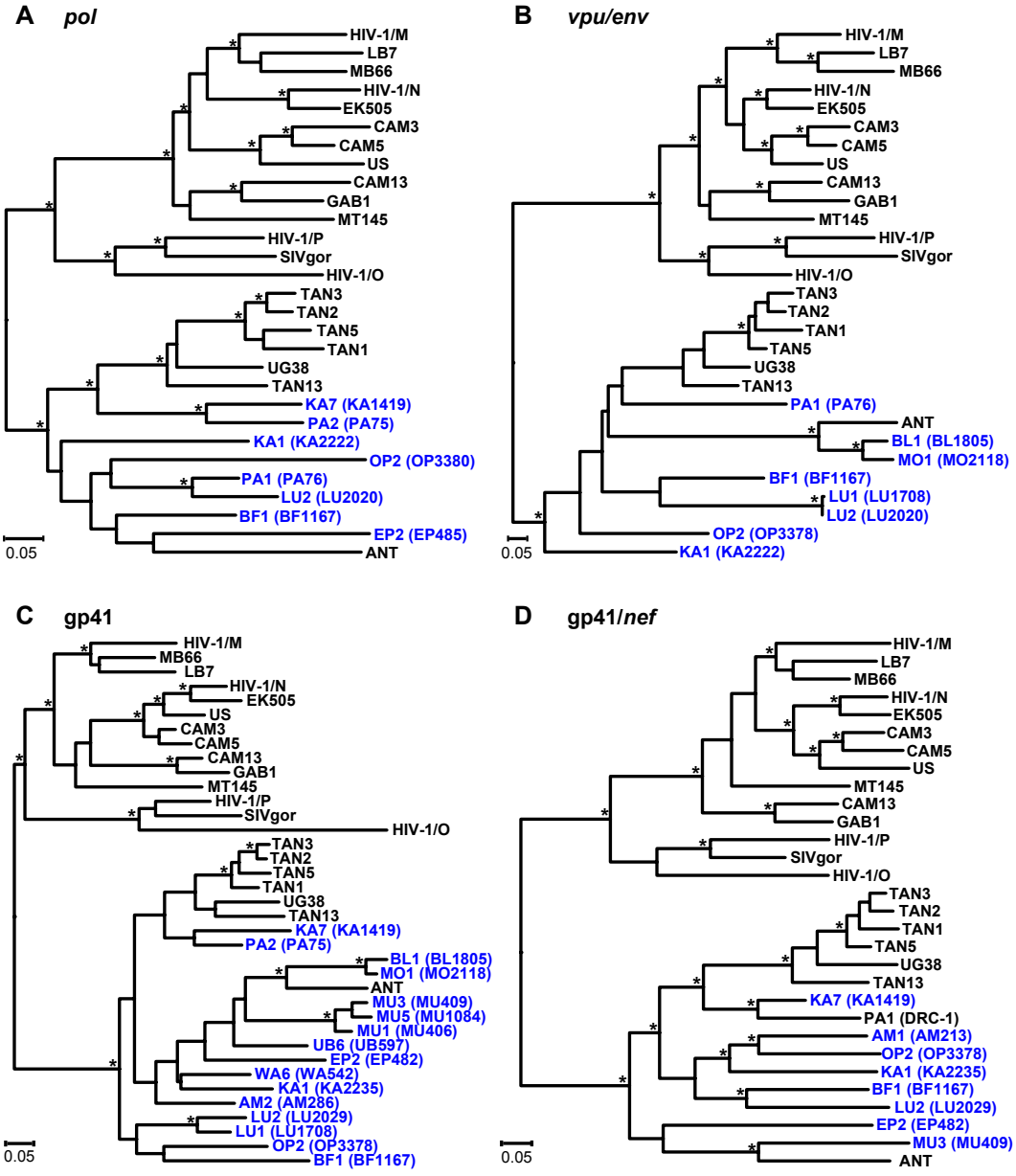
A pol
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
CAM13
GAB1
MT145
HIV-1/P
SIVgor
HIV-1/O
TAN3
TAN2
TAN5
TAN1
UG38
TAN13
KA7 (KA1419)
PA2 (PA75)
KA1 (KA2222)
OP2 (OP3380)
PA1 (PA76)
LU2 (LU2020)
BF1 (BF1167)
EP2 (EP485)
ANT
0.05
B vpu/env
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
CAM13
GAB1
MT145
HIV-1/P
SIVgor
HIV-1/O
TAN3
TAN2
TAN1
TAN5
UG38
TAN13
PA1 (PA76)
ANT
BL1 (BL1805)
MO1 (MO2118)
BF1 (BF1167)
LU1 (LU1708)
LU2 (LU2020)
OP2 (OP3378)
KA1 (KA2222)
0.05
C gp41
HIV-1/M
MB66
LB7
HIV-1/N
EK505
US
CAM3
CAM5
CAM13
GAB1
MT145
HIV-1/P
SIVgor
HIV-1/O
TAN3
TAN2
TAN5
TAN1
UG38
TAN13
KA7 (KA1419)
PA2 (PA75)
BL1 (BL1805)
MO1 (MO2118)
ANT
MU3 (MU409)
MU5 (MU1084)
MU1 (MU406)
UB6 (UB597)
EP2 (EP482)
WA6 (WA542)
KA1 (KA2235)
AM2 (AM286)
LU2 (LU2029)
LU1 (LU1708)
OP2 (OP3378)
BF1 (BF1167)
0.05
D gp41/nef
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
MT145
CAM13
GAB1
HIV-1/P
SIVgor
HIV-1/O
TAN3
TAN2
TAN1
TAN5
UG38
TAN13
KA7 (KA1419)
PA1 (DRC-1)
AM1 (AM213)
OP2 (OP3378)
KA1 (KA2235)
BF1 (BF1167)
LU2 (LU2029)
EP2 (EP482)
MU3 (MU409)
ANT
0.05

**Figure 4**

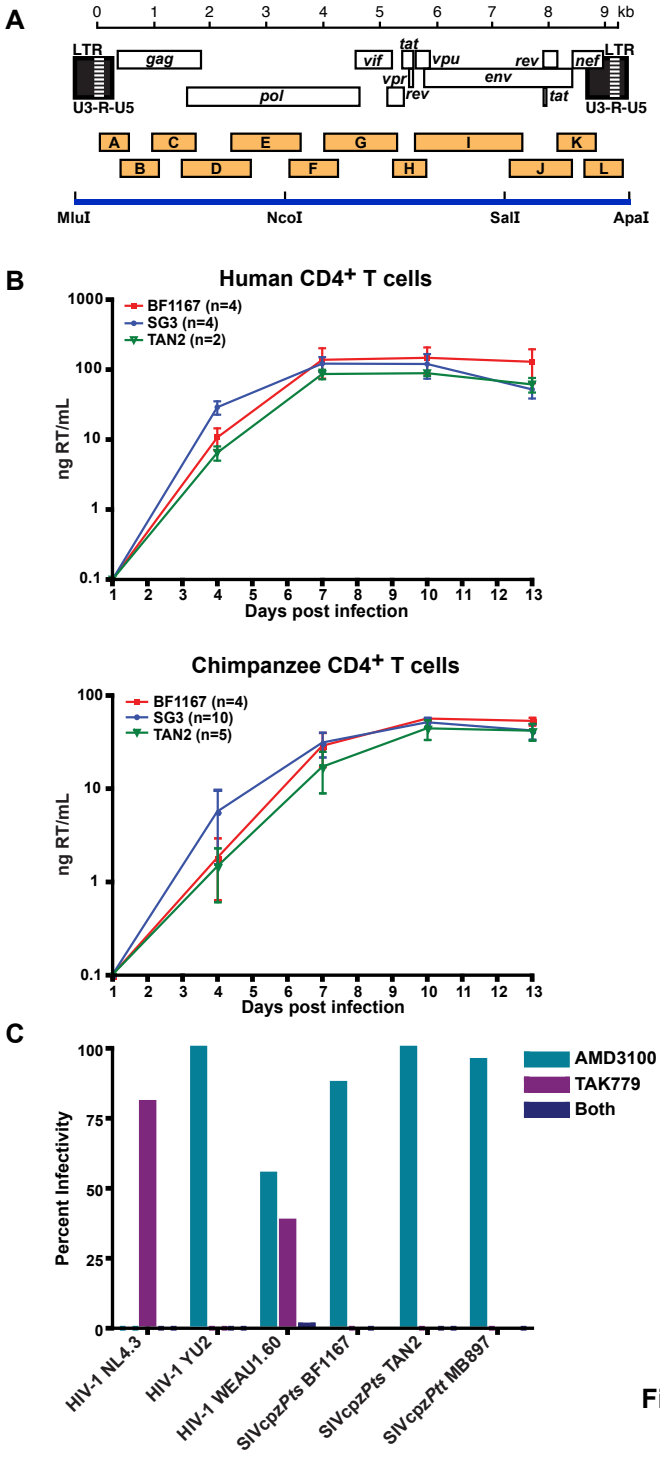
A
0 1 2 3 4 5 6 7 8 9 kb
LTR
gag
pol
vif
tat
vpu
rev
vpr
env
rev
tat
nef
LTR
U3-R-U5
U3-R-U5
A C E G I K
B D F H J L
MluI
NcoI
SalI
ApaI
B
Human CD4+ T cells
BF1167 (n=4)
SG3 (n=4)
TAN2 (n=2)
ng RT/mL
Days post infection
Chimpanzee CD4+ T cells
BF1167 (n=4)
SG3 (n=10)
TAN2 (n=5)
ng RT/mL
Days post infection
C
Percent Infectivity
AMD3100
TAK779
Both
HIV-1 NL4.3
HIV-1 YU2
HIV-1 WEAU1.60
SIVcpzPts BF1167
SIVcpzPts TAN2
SIVcpzPtt MB897

Figure 5

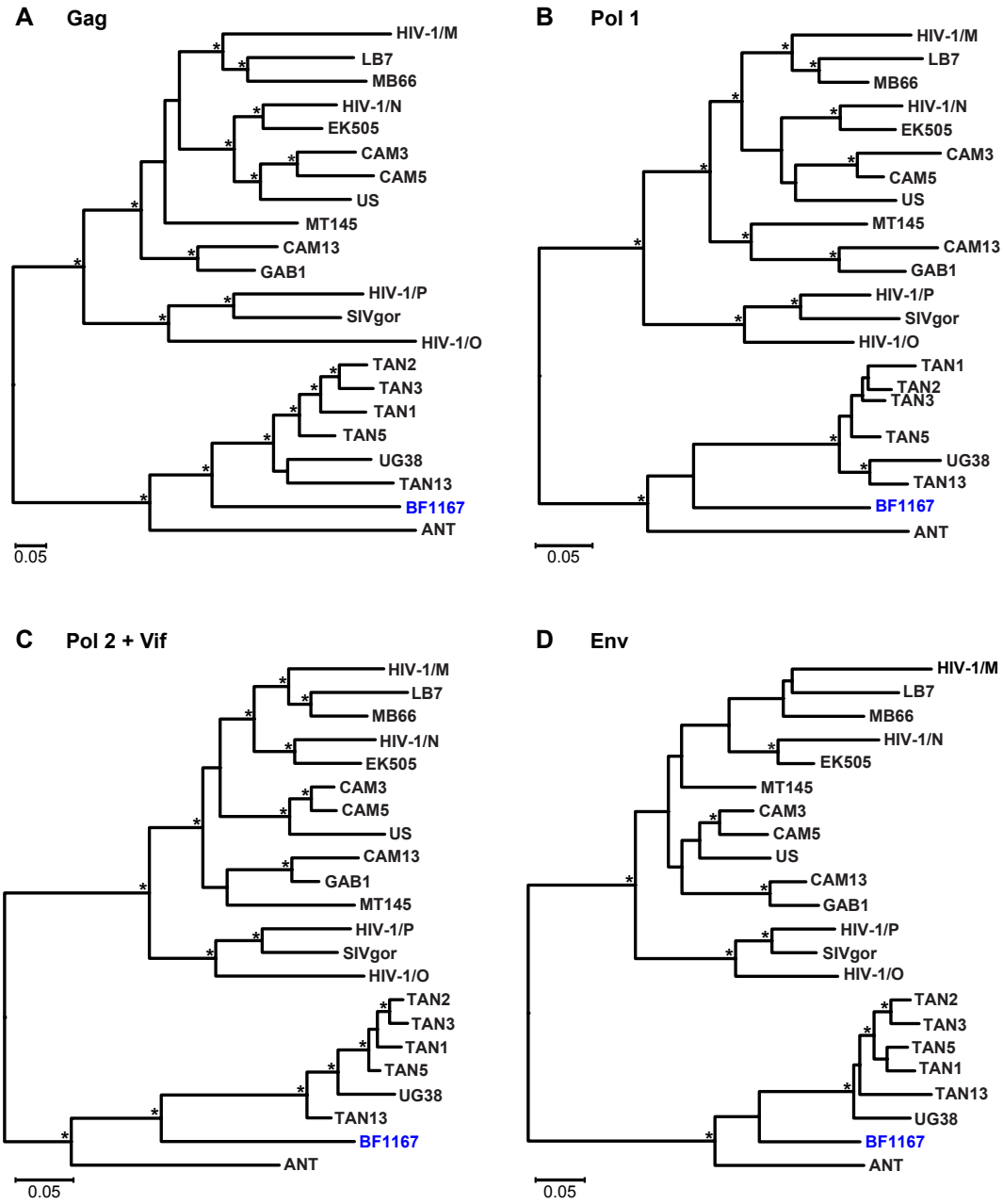
A
Gag
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
MT145
CAM13
GAB1
HIV-1/P
SIVgor
HIV-1/O
TAN2
TAN3
TAN1
TAN5
UG38
TAN13
BF1167
ANT
0.05
B
Pol 1
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
MT145
CAM13
GAB1
HIV-1/P
SIVgor
HIV-1/O
TAN1
TAN2
TAN3
TAN5
UG38
TAN13
BF1167
ANT
0.05
C
Pol 2 + Vif
HIV-1/M
LB7
MB66
HIV-1/N
EK505
CAM3
CAM5
US
CAM13
GAB1
MT145
HIV-1/P
SIVgor
HIV-1/O
TAN2
TAN3
TAN1
TAN5
UG38
TAN13
BF1167
ANT
0.05
D
Env
HIV-1/M
LB7
MB66
HIV-1/N
EK505
MT145
CAM3
CAM5
US
CAM13
GAB1
HIV-1/P
SIVgor
HIV-1/O
TAN2
TAN3
TAN5
TAN1
TAN13
UG38
BF1167
ANT
0.05

**Figure 6**

| Strain | Gag-30 | Arg | Lys |
|---|---|---|---|
| SIVcpz*Pts* | | | |
| BF1167 | Leu CTT | 1 | 3 |
| KA1 | Leu TTA | 2 | 2 |
| LU2 | Leu TTG | 2 | 2 |
| MO1 | Met ATG | 1 | 1 |
| OP2 | Leu TTG | 2 | 2 |
| PA1 | Leu TTG | 2 | 2 |
| UB6 | Met ATG | 1 | 1 |
| ANT | Met ATG | 1 | 1 |
| TAN1 | Leu TTA | 2 | 2 |
| TAN2 | Leu TTA | 2 | 2 |
| TAN3 | Leu TTA | 2 | 2 |
| TAN5 | Leu TTG | 2 | 2 |
| TAN13 | Leu TTA | 2 | 2 |
| UG38 | Leu TTG | 2 | 2 |
| SIVcpz*Ptt* ($n$ = 15) | Met ATG | 1 | 1 |
| SIVgor ($n$ = 3) | Met ATG | 1 | 1 |

**Figure 7**

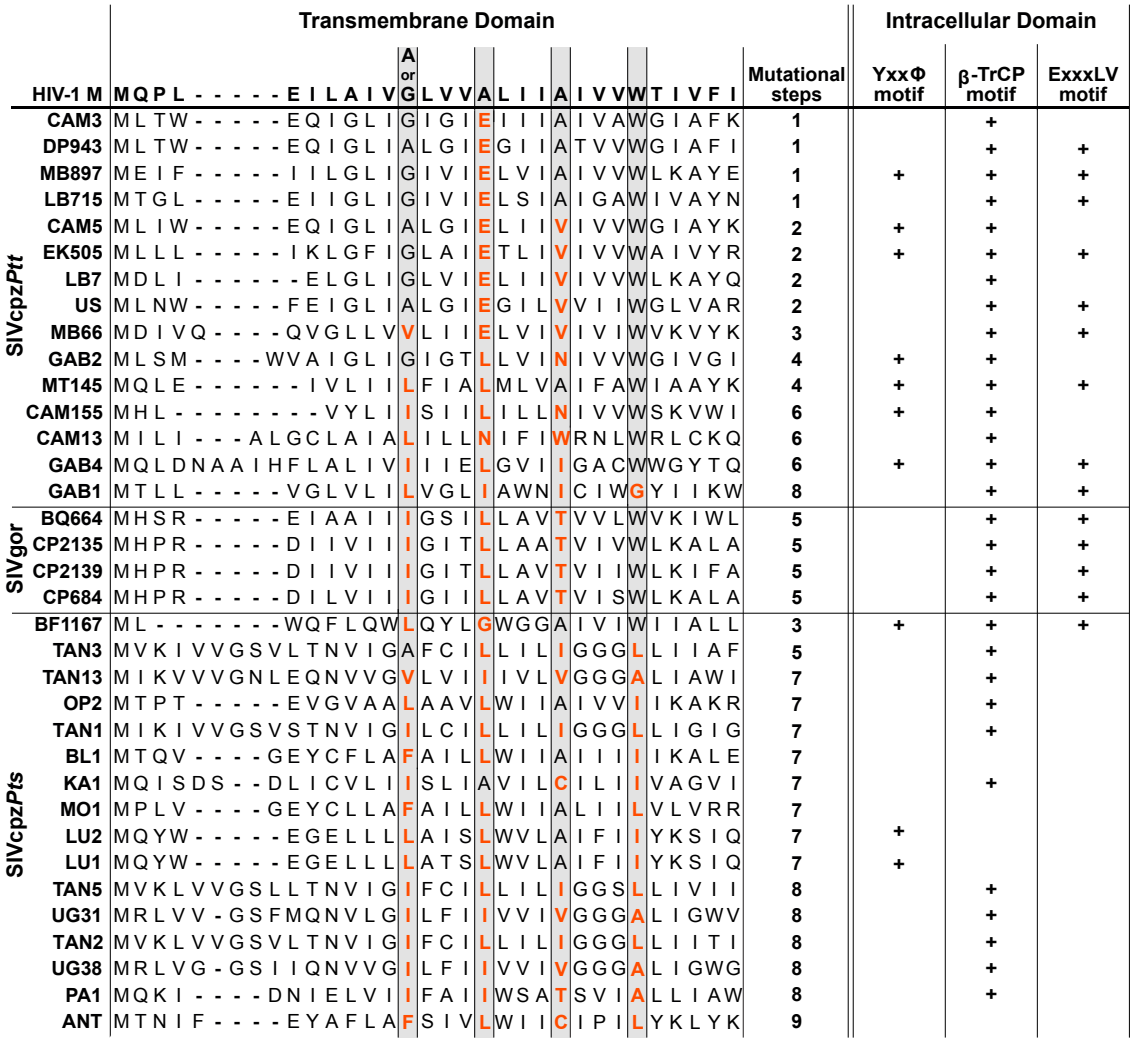
Transmembrane Domain
Intracellular Domain
HIV-1 M MQPL-----EILAIVAorGLVVALIIAIVVWTIVFI
Mutational steps
YxxΦ motif
β-TrCP motif
ExxxLV motif
SIVcpzPtt
CAM3 MLTW-----EQIGLIGIGIEIIIAIVAWGIAFK 1 + +
DP943 MLTW-----EQIGLIALGIEGIIATVVWGIAFI 1 + +
MB897 MEIF-----IILGLIGIVIELVIAIVVWLKAYE 1 + + +
LB715 MTGL-----EIIGLIGIVIELSIAIGAWIVAYN 1 + +
CAM5 MLIW-----EQIGLIALGIELIIVIVVWGIAYK 2 + +
EK505 MLLL-----IKLGFIGLAIETLIVIVVWAIVYR 2 + + +
LB7 MDLI-------ELGLIGLVIELIIVIVVWLKAYQ 2 +
US MLNW-----FEIGLIALGIEGILVVIIWGLVAR 2 + +
MB66 MDIVQ----QVGLLVVLIIELVIVIVIWVKVYK 3 + +
GAB2 MLSM----WVAIGLIGIGTLLVINIVVWGIVGI 4 + +
MT145 MQLE-------IVLIILFIALMLVAIFAWIAAYK 4 + + +
CAM155 MHL---------VYLIISIILLNIVVWSKVWI 6 + +
CAM13 MILI---ALGCLAIALILLNIFIWRNLWRLCKQ 6 +
GAB4 MQLDNAAIHFLALIVIIIELGVIIGACWWGYTQ 6 + + +
GAB1 MTLL-----VGLVLILVGLIAWNICIWGYIIKW 8 + +
SIVgor
BQ664 MHSR-----EIAAIIIGSILLAVTVVLWVKIWL 5 + +
CP2135 MHPR-----DIIVIIIGITLLAATVIVWLKALA 5 + +
CP2139 MHPR-----DIIVIIIGITLLAVTVIIWLKIFA 5 + +
CP684 MHPR-----DILVIIIGIILLAVTVISWLKALA 5 + +
SIVcpzPts
BF1167 ML-------WQFLQWLQYLGWGGAIVIWIIALL 3 + + +
TAN3 MVKIVVGSVLTNVIGAFCILLILIGGGLLIIAF 5 +
TAN13 MIKVVVGNLEQNVVGVLVIIIVLVGGGALIAWI 7 +
OP2 MTPT-----EVGVAALAAVLWIIAIVVIIKAKR 7 +
TAN1 MIKIVVGSVSTNVIGILCILLILIGGGLLIGIG 7 +
BL1 MTQV----GEYCFLAFAILLWIIAIIIIIKALE 7
KA1 MQISDS--DLICVLIISLIAVILCILIIVAGVI 7 +
MO1 MPLV----GEYCLLAFAILLWIIALIILVLVRR 7
LU2 MQYW-----EGELLLLAISLWVLAIFIIYKSIQ 7 +
LU1 MQYW-----EGELLLLATSLWVLAIFIIYKSIQ 7 +
TAN5 MVKLVVGSLLTNVIGIFCILLILIGGSLLIVII 8 +
UG31 MRLVV-GSFMQNVLGILFIIVVIVGGGALIGWV 8 +
TAN2 MVKLVVGSVLTNVIGIFCILLILIGGGLLIITI 8 +
UG38 MRLVG-GSIIQNVVGILFIIVVIVGGGALIGWG 8 +
PA1 MQKI----DNIELVIIFAIIWSATSVIALLIAW 8 +
ANT MTNIF----EYAFLAFSIVLWIICIPILYKLYK 9

**Figure 8**