SUPPLEMENTARY INFORMATION

Supplementary Methods

Sample collection. To extend our previous molecular epidemiological survey of SIV infection in wild-living ape populations[1], 725 additional fecal samples were collected at 15 remote forest sites in southern Cameroon (Supplementary Table 1). These included 378 samples from chimpanzees (*Pan troglodytes*) and 213 samples from gorillas (*Gorilla gorilla*) (the remainder were found to be degraded or from other primate/mammal species following mitochondrial DNA analysis; see below). Gorilla samples were identified at 11 forest sites (Supplementary Fig. 1). One of these (TK) was located in the range of the Cross River Gorilla[2] (*G. g. diehli*); all other sites were within the range of the Western Lowland Gorilla[2] (*G. g. gorilla*) in either protected (CP, EK, BM, LB, BB, NK) or unprotected areas (DP, BQ, KG, MB). Samples were collected by experienced trackers with logistical support from Project PRESICA, Project Grand Singes, the World Wildlife Fund and the Cameroonian Ministry of Environment and Forestry. Collection site, date, global positioning system (GPS) coordinates (when available), and species origin (by visual inspection) were recorded. Specimens were preserved in RNAlater (Ambion, Austin, TX), kept at base camps at ambient temperature for up to 20 days, and subsequently transported to a central laboratory in Yaounde for storage at –20°C as described[1]. Samples were shipped to Europe and the US at ambient temperatures, and stored at –80°C upon receipt.

Detection of HIV-1 cross-reactive antibodies in gorilla fecal samples. Fecal extracts suitable for immunoblot analysis were prepared as described[1]. Because the sensitivity and specificity of SIV antibody detection in gorilla fecal samples are

unknown, all gorilla samples were examined by two different immunoblot assays. The INNO-LIA HIV Confirmation test (Innogenetics, Ghent, Belgium) contains HIV-1 and HIV-2 recombinant proteins and synthetic peptides coated as discrete lines on a nylon strip (Supplementary Fig. 2) and has been used in the past to identify SIV infection in a number of different primate species[3]. The enhanced chemiluminescent Western blot assay (Calypte Biomedical; Rockville, MD) is a confirmatory test for HIV-1 infection in humans and detects SIVcpz antibodies in fecal samples from wild chimpanzees with 92% sensitivity and 100% specificity[1,4]. Six of 213 gorilla fecal samples scored positive in both assays, reacting with the HIV-1 gp41 protein on the INNO-LIA (Supplementary Fig. 2), and HIV-1 Env (gp160, gp120 and gp41) and Pol (p66, p55, p31) proteins on Western blot strips (Fig. 1a); surprisingly, no reactivity with HIV-1 Gag (p24, p17) proteins was detected. On Western blots, one sample reacted faintly with HIV-1 gp120 and another with HIV-1 p24, but both were INNO-LIA negative. Since attempts to amplify viral sequences were repeatedly negative (not shown), their reactivity was likely unspecific. All other samples were negative in both assays.

Amplification of SIVgor sequences from fecal RNA. Fecal RNA was extracted and subjected to reverse transcription polymerase chain reaction (RT-PCR) amplification as described[1]. For all amplifications, cDNA was synthesized using the outer reverse primer (R1), followed by nested PCR using degenerate HIV-1/SIVcpz consensus or strain specific primer pairs. Initially, a ~340 bp diagnostic SIVgor *pol* fragment was amplified from four antibody positive gorilla samples using primers F1/R1 followed by F2/R2 (F1: 5'-CCAGCNCACAAAGGNATAGGAGG-3'; R1: 5'-

ACBACYGCNCCTTCHCCTTTC-3'; F2: 5'-GGAAGTGGATACTTAGAAGCAGAAGT-3'; R2: 5'-CCAATYCCYCCYYTTYKYTTAAAATT-3'). A larger ~940 bp *pol* fragment was subsequently amplified from three of these same samples using a combination of two forward (F1a: 5'-ACCTGGATNCCWGANTGGGA-3' plus F1b: 5'-TGGTGGHCWGABTAYTGGCA-3') and two reverse (R1a: 5'-ACBACYGCNCCTTCHCCTTTC-3' plus R1b: 5'-ACTGCHCCYTCWCCTTTCCACAG-3') primers in the first round, followed by single primers (F2: 5'-TWYTATGTWGATGGRGCAGC-3'; R2: 5'-CCAATYCCYCCYYTTYKYTTAAAATT-3') in the second round. For sample CP684, an additional 770 bp fragment was amplified using consensus forward and strain specific reverse primers (F1: 5'-AATAARAGRACHCARGANTTYTGGGA-3'; R1: 5'-CCCCAGCACATACTGGGAG-3'; F2: 5'-CCHCCNTTYYTNTGGATGGG-3'; R2: 5'-CTTCCAAGGCCAAGAGGACTG-3'). Together with the 940 bp *pol* fragment, this yielded 1,537 bp of *pol* sequence for CP684 (Supplementary Table 2). For four samples, a ~440 bp fragment spanning the gp41 ectodomain was amplified as described[1] with some primer modifications (F1: 5'-GCWGGHASYACWATGGGCGCAG-3'; R1: 5'-AGRGGWKKATAWCCCTGCCTAA-3'; F2: 5'-DCTGRKATWGTRCARCAGCA-3' and R2: 5'-TCCTACTATSATWATARMTATTTTTATATA-3'). Finally, for sample CP1436 a ~790bp fragment spanning the SIVgor gp41/nef junction was amplified (F1: 5'-AWTGGYTRWGGTAYATHARRAT-3'; R1: 5'-CCCHTCCAGTCCYCCCTTTTC-3'; F2a: 5'-GCTTAAGAAAGGTTAGGCAGGG-3' plus F2b: 5'-TAGTAAMAAAAGTTAGGCAGGG-3'; R2: 5'-CAGTCCYCCCTTTTCTTTYAAAAA-3'). RT-PCR products were gel purified and sequenced directly (GenBank accession numbers AM296484 to AM296492).

Phylogenetic analyses of SIVgor sequences. Deduced amino acid sequences of SIVgor gene fragments were aligned with corresponding sequences of representative SIVcpz and HIV-1 strains using ClustalW[5] (GenBank accession numbers were as follows: HIV-1 M/A U455, M62320; HIV-1 M/B HXB2, K03455; HIV-1 N YBF30, AJ006022; HIV-1 N YBF106, AJ271370; HIV-1 O ANT70, L20587; HIV-1 O MVP5180, L20571; SIVcpzGAB1, X52154; SIVcpzGAB2, AF382828; SIVcpzCAM3, AF115393; SIVcpzCAM5, AJ271369; SIVcpzCAM13, AY169968; SIVcpzUS, AF103818; SIVcpzMB66, DQ373063; SIVcpzLB7, DQ373064; SIVcpzEK505, DQ373065; SIVcpzMT145, DQ373066; SIVcpzTAN1, AF447763; SIVcpzTAN2, DQ374657; SIVcpzTAN3, DQ374658; SIVcpzANT, U42720; SIVcpzMT115, DQ370395 and DQ370370; SIVcpzDP206, DQ370403 and DQ370375; SIVcpzDP25, DQ370405 and DQ370378; SIVcpzEK502, DQ370408 and DQ370381; SIVcpzEK516, DQ370382; SIVcpzEK519, DQ370411 and DQ370383; SIVcpzMB23, DQ370413 and DQ370388; SIVcpzMB192, DQ370415 and DQ370392; SIVcpzMB317, DQ370416 and DQ370387; SIVcpzMB245, DQ370417 and DQ370385; SIVcpzLB186, DQ370418; and SIVcpzMB248, DQ370386). Sites that could not be unambiguously aligned and sites with a gap in any sequence were excluded. Trees were inferred by the Bayesian method[6], implemented in MrBayes (v3.1.2)[7,8], using the Jones, Taylor and Thornton matrix[9] and gamma-distributed rates at sites[10], with one million generations and burn-in of 25%. Bayesian parameters were examined with the Tracer program and all estimated sample sizes were greater than 779 (http://evolve.zoo.ox.ac.uk/software.html/id=tracer).

Species and subspecies determinations. The species origin of the chimpanzee and gorilla samples was determined by mitochondrial DNA (mtDNA) analysis. First, a ~450 to 500 bp mtDNA fragment spanning the hypervariable D loop was amplified from fecal DNA using primers L15997 (5'-CACCATTAGCACCCAAAGCT-3') and H16498 (5'-CCTGAAGTAGGAACCAGATG-3') as described[2]. Phylogenetic analysis of these D loop sequences permitted classification of all chimpanzee (including subspecies analysis; Supplementary Table 1) and the majority of gorilla samples (n=131). Gorilla samples that yielded D loop sequences of poor quality (n=82) were reanalyzed by amplifying a 386 bp mtDNA fragment spanning the 12SrRNA gene[11] (using primers 12S-L1091 5'-AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT-3' and 12S-H1478 5'-TGACTGCAGAGGGTGACGGGCGGTGTGT-3'). Moreover, a subset (n=51) of gorilla samples, including all six SIVgor positive specimen, was also subjected to a nested PCR strategy which targeted a ~7,400 bp mtDNA fragment in the first (F1 5'-GACTTCTGGCAAGCCTCGCCAA-3'; R1 5'-ATTATACTCCGAGGTCGC-3'), and a ~1,000 bp mtDNA fragment in the second round (F2 5'-CCACATCAAACCCGAATGATA-3'; R2 5'-TCAAAGACAGATACTGCGACAT-3') of PCR. The rationale for this approach was to reduce the number of mtDNA sequences that comprised nuclear inserts (numts) which frequently confound phylogenetic analyses of gorilla mtDNA sequences[12]. This approach generated 47 genuine mtDNA and 4 numt sequences which were identified by comparison to reference sequences in the database[12,13] (not shown). The 47 genuine sequences represented 23 different haplotypes (GenBank accession numbers AM392403 to AM392425; Supplementary Table 3) which were aligned with corresponding regions of fully

sequenced mitochondrial genomes from western gorillas (*G. gorilla*) (GenBank accession number D38114 and X93347), humans (AF381992 and AF346975), chimpanzees (*Pan troglodytes*) (D38113, X93335), a bonobo (*Pan paniscus*) (D38116) and orangutans (*Pongo pygmaeus*) (D38115, X97707) using ClustalW[5], resulting in a gap-stripped alignment of 939 nucleotides. A phylogenetic tree was constructed using the Bayesian method[6], implemented in MrBayes (v3.1.2)[7,8] using the HKY model[14] for gamma-distributed rates at sites[10] and 10 million generations with a burn-in of 25% (Supplementary Fig. 3). Estimated sample sizes from the Tracer program (http://evolve.zoo.ox.ac.uk/software.html/id=tracer) were greater than 12,000. Finally, to exclude any admixture of chimpanzee feces, all SIVgor positive fecal samples were subjected to mtDNA amplification using "chimpanzee specific" primer pairs. Three different sets were designed, in each case placing one of the primers in a region of the mtDNA genome where gorillas lack a sequence segment that is retained in humans and chimpanzees (set 1: F1 primer L15997[2]; R1a 5'-GACAGTTGRGGGTTRATTGTTGT-3' plus R1b 5'-TGTTGTRCRTGCTTGTAAGCAT-3'; set 2: F2a 5'-ACAACAATYAACCCYCAACTGTC-3' plus F2b 5'-ACATAANACRCAACYCCAA-3'; R2 primer H16498[2]; set 3: F3 5'-GACATCACGATGGATCACAGGTC-3'; R3a 5'-GGGGCTAGTAGAATGGGAGTT-3' or R3b 5'-TCTGGCTAGGCTGGTGTTTGGG-3'). None of the SIVgor positive fecal samples, but three of three chimpanzee controls, yielded amplification products with these primer sets (not shown).

Microsatellite and gender analyses. Microsatellite analysis was performed essentially as described[1]. Samples were genotyped at six loci (D18s536, D4s243,

D2s1326, D2s1333, D4s1627, D9s905) with all PCR reactions performed in duplicate. Individuals whose genotype appeared homozygous were amplified a minimum of seven times to exclude allelic drop out. For gender determination, a region of the amelogenin gene was amplified using primers AMEL-F212 and AMEL-R212 as described[1]. The amelogenin gene contains a 6 bp deletion in the X, but not the Y chromosome, resulting in amplicons of 212 bp and 218 bp for X and Y chromosomes, respectively[15]. Products were visualized and sized as the microsatellite amplicons. The results of microsatellite analysis are shown in Supplementary Table 2.

References

1. Keele, B. F. *et al*. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523-526 (2006).
2. Butynski, T. M. in *Great Apes and Humans: The Ethics of Coexistence* (eds. Beck, B. B. et al.) 3-56 (Smithsonian Institution Press, Washington and London, 2001).
3. Peeters M. *et al.,* Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis.* **8**, 451-457 (2002).
4. Santiago, M.L., *et al*. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J. Virol*. **77**, 7545-7562 (2003).
5. Thompson, J. D., Higgins, D. G., & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
6. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754-755 (2003).
7. Ronquist, F. & Huelsenbeck, J.P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**,1572-1574 (2003).
8. Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, & Ronquist, F. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407-415 (2004).
9. Jones, D.T., Taylor, W.R., & Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci*. 8, 275-282 (1992).
10. Yang, Z. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol*. 10,1396-1401 (1993).
11. van der Kuyl, A. C., Dekker, J. T. & Goudsmit, J. Primate genus *Miopithecus*: evidence for the existence of species and subspecies of dwarf guenons based on cellular and endogenous viral sequences. *Mol. Phylogenet. Evol*. 14, 403-413 (2000).
12. Thalmann, O. *et al*. Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Mol. Ecol.* 14, 179-188 (2005).

13. Clifford, S. L. *et al*. Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). Mol. Ecol. **13**, 1551-1565 (2004).
14. Hasegawa, M., Kishino, H. & Yano, T. Dating the human-ape split by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**,160-174 (1985).
15. Sullivan, K. M., Mannucci, A., Kimpton, C. P. & Gill, P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* **15**, 636-638, 640-641 (1993).

Supplementary Figure 1

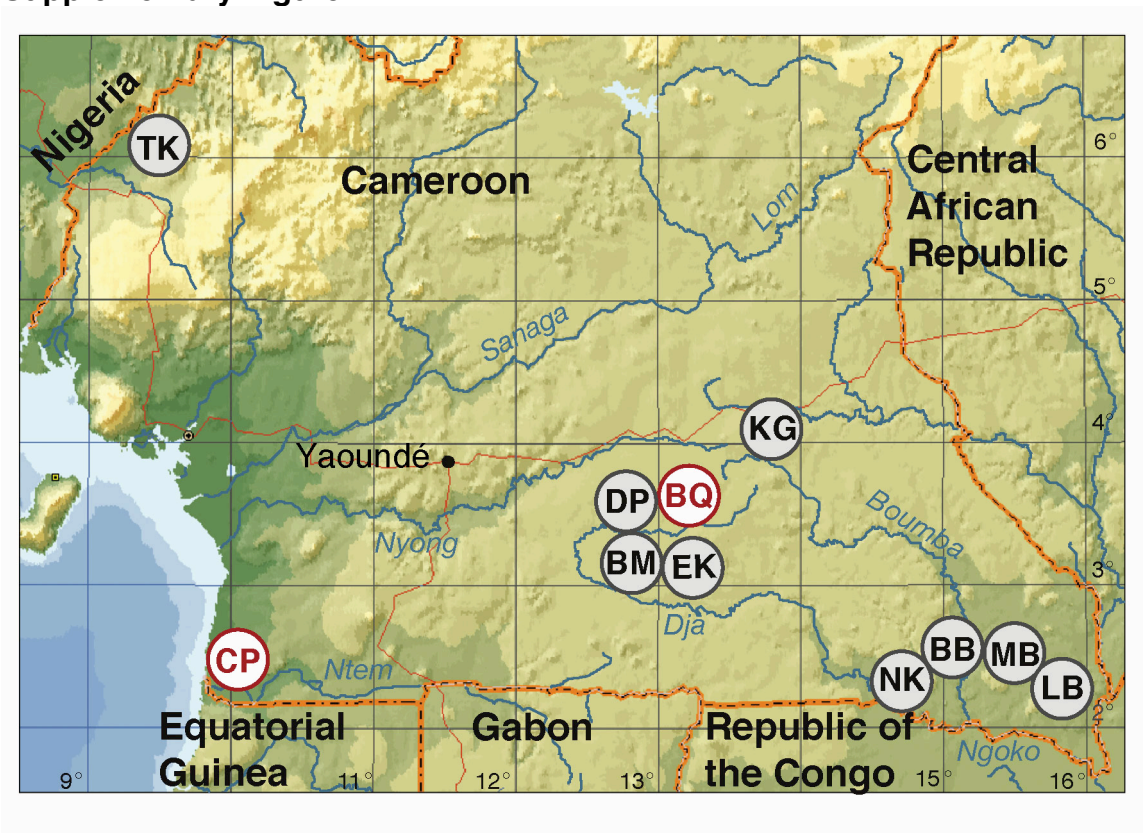

Fig. S1. Forest sites in Cameroon from which gorilla fecal samples were collected. Sampling sites (circled) are indicated by two-letter initials. Sites where SIVgor infection was found are highlighted in red.

Supplementary Figure 2

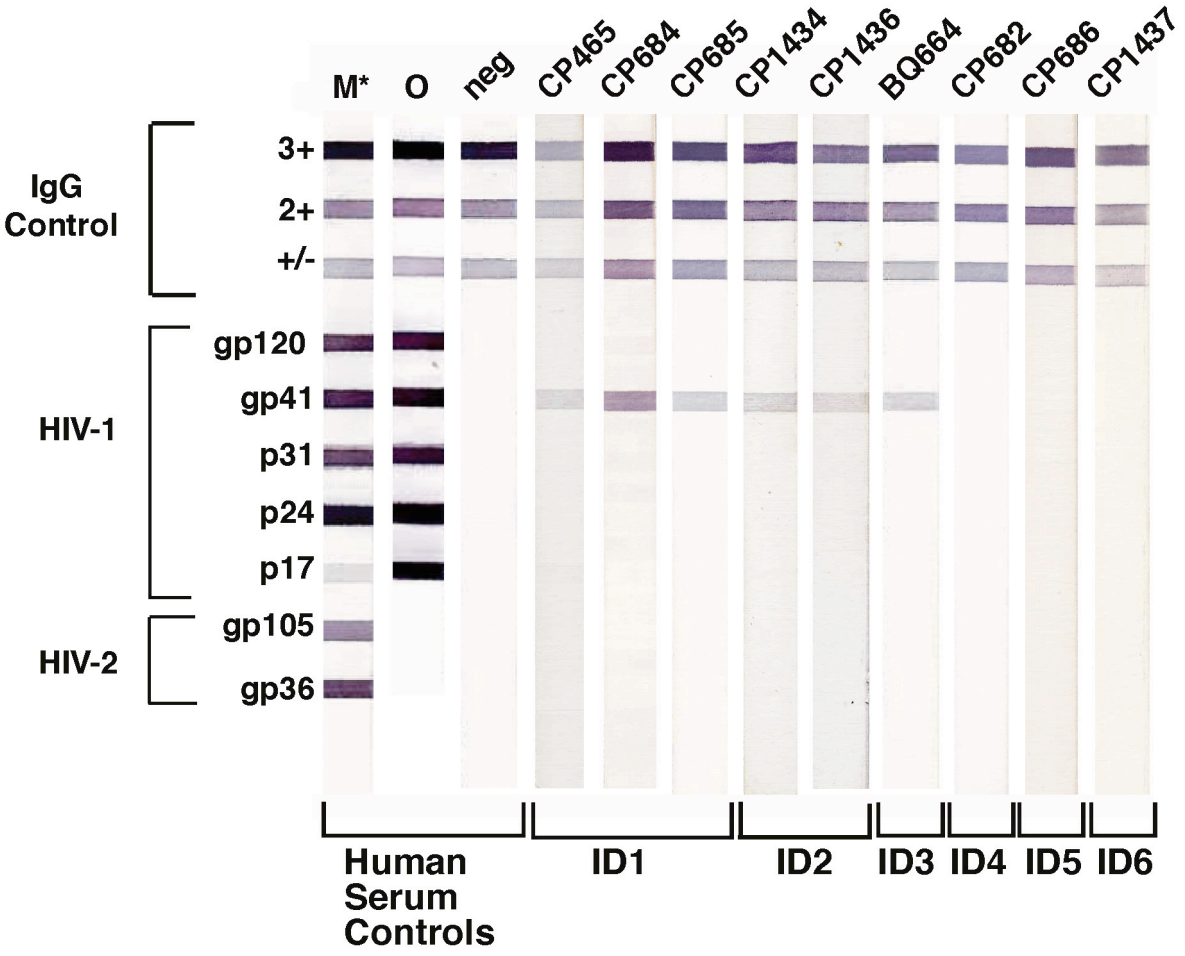

Fig. S2. Detection of HIV-1 cross-reactive antibodies in gorilla fecal samples using the INNO-LIA HIV Confirmation test. HIV-1 and HIV-2 recombinant proteins and synthetic peptides are coated as indicated (the 3+, 2+ and +/- bands on the top represent control lanes for sample addition and test performance). Fecal extracts are numbered with letters indicating their collection site as shown in Supplementary Fig. 1. Samples from the same individual (ID) are grouped (Supplementary Table 2). Fecal samples from three gorillas (ID1-ID3) were found to recognize at least one HIV-1 antigen (gp41) with a band intensity equal or greater than the assay cut-off (+/-), and were thus scored as SIV antibody positive; all other fecal samples (including those of gorillas ID4-ID6) exhibited no reactivity and were thus scored as SIV antibody negative. The banding patterns of a dually reactive (HIV-1 group M plus HIV-2) serum included in the INNO-LIA kit (M*), as well as those of sera from HIV-1 group O infected (O) and uninfected (neg) humans, are shown for control.

Supplementary Figure 3

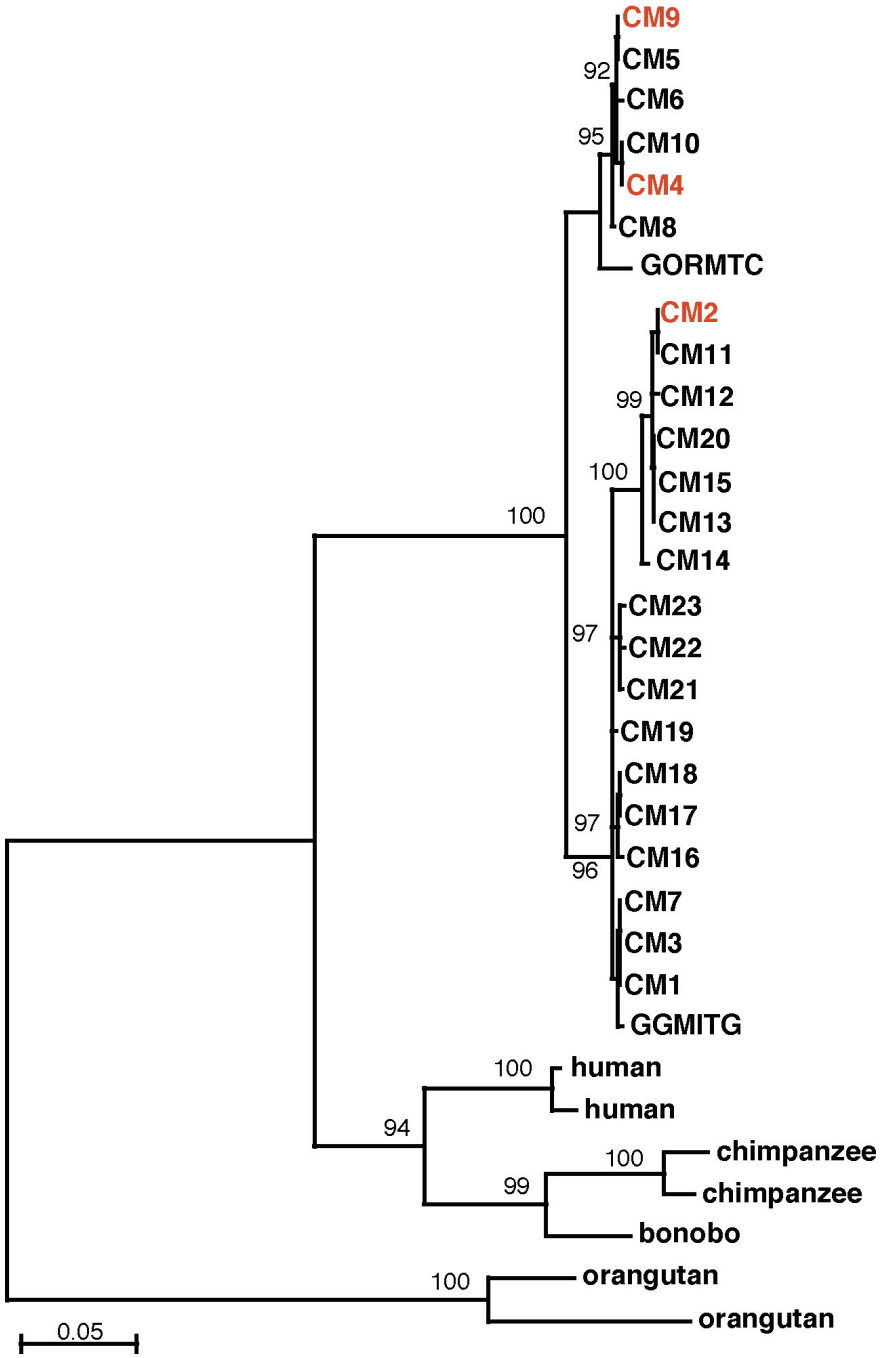

Fig. S3. Species origin of SIVgor positive fecal samples. Mitochondrial DNA sequences from 47 gorilla fecal samples (consensus length 939 bp) were grouped into unique haplotypes (Supplementary Table 3) and compared to species specific mtDNA reference sequences from the database (GORMTC and GGMITG are both derived from western gorillas). Haplotypes of SIVgor infected gorillas are shown in red. The phylogenetic tree was inferred by the Bayesian method. Numbers on nodes are estimated posterior probabilities of 90% and above. The scale bar represents 0.05 substitutions per site.

Supplementary Figure 4

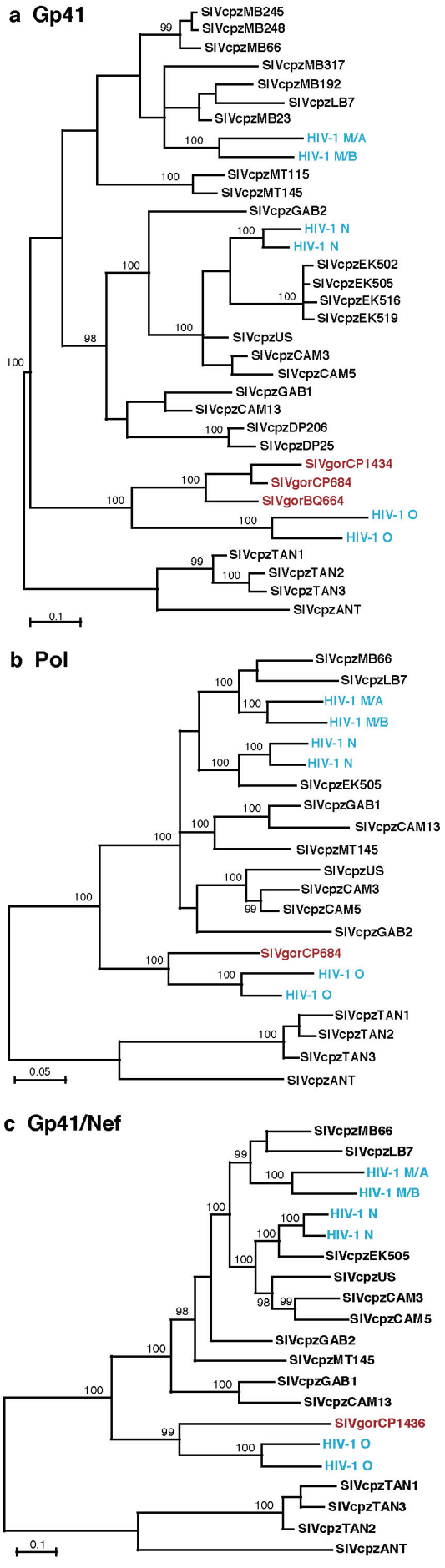

Fig. S4. Evolutionary relationship of SIVgor from wild-living gorillas to SIVcpz and HIV-1 strains. Trees are inferred from partial Gp41 (a), Pol (b) and Gp41/Nef (c) protein sequences by the Bayesian method (consensus lengths 122, 510, and 216 amino acids, respectively). Newly identified SIVgor strains are shown in red; the three groups of HIV-1 are highlighted in blue. SIVcpz strains TAN1, TAN2, TAN3 and ANT are from eastern chimpanzees (*P. t. schweinfurthii*); all other SIVcpz strains are from central chimpanzees (*P. t. troglodytes*). Numbers on internal branches are estimated posterior probabilities of 95% and above. Scale bars represent 0.1 and 0.05 substitutions per site.

Supplementary Table 1

Table S1. SIV infection in wild-living chimpanzee and gorilla communities

Collection Sites[1]	Chimpanzee samples[2]	*P.t.troglodytes* samples[2]	*P.t.vellerosus* samples[2]	*P. t. troglodytes* samples with HIV-1 crossreactive antibodies	Gorilla samples[2]	*G. gorilla* samples with HIV-1 crossreactive antibodies	Number of SIVgor infected gorillas[3]
TK	1	0	1	0	21	0	0
MF	39	0	39	0	0	0	0
CP	42	42	0	0	51	5	2
MP	15	0	15	0	0	0	0
MG	24	24	0	0	0	0	0
KG	15	15	0	0	15	0	0
DP	35	35	0	10	16	0	0
BQ	10	10	0	0	25	1	1
SL	44	44	0	5	0	0	0
EK	27	27	0	0	6	0	0
BM	38	38	0	2	1	0	0
LB	18	18	0	3	48	0	0
MB	62	62	0	20	2	0	0
NK	8	8	0	0	25	0	0
BB	0	0	0	0	3	0	0
Total	**378**	**323**	**55**	**40**	**213**	**6**	**3**

[1]Location of sites is shown in Supplementary Fig. 1, except for MF (40 km south of TK), MP (185 km northwest of KG; north of the Sanaga River), MG (138 km northwest of KG; south of the Sanaga River) and SL (36 km southwest of DP; north of the Dja River).

[2]as determined by mitochondrial DNA analysis of fecal DNA (see methods).

[3]as determined by microsatellite analysis (Supplementary Table 2).

Supplementary Table 2

Table S2. Mitochondrial and microsatellite analysis of SIVgor positive fecal samples from wild-living gorillas.

					Amplification of SIVgor virion RNA (bp)										
gorilla	Fecal sample	Date of sample collection	Fecal INNOLIA analysis[1]	Fecal Western blot analysis[2]	*pol*	*gp41*	*gp41/ nef*	mtDNA haplotype[3]	Sex[4]	Locus[5] D18s536	Locus[5] D4s243	Locus[5] D2s1326	Locus[5] D2s1333	Locus[5] D4s1627	Locus[5] D9s905
ID1	CP465[6]	04/18/04	Pos	Pos	n/a[7]	n/a	n/a	CM4	F	145/149	n/a	n/a	n/a	236/236	275/278
	CP684	04/18/04	Pos	Pos	1,537	394	n/a	CM4	F	145/149	187/195	251/251	298/317	236/236	275/278
	CP685	04/19/04	Pos	Pos	280	394	n/a	CM4	F	145/149	187/195	251/251	298/317	236/236	275/278
ID2	CP1434	03/31/06	Pos	Pos	n/a	388	n/a	CM9	F	149/157	183/195	255/267	295/295	235/244	278/278
	CP1436	03/31/06	Pos	Pos	892	n/a	738	CM9	F	149/157	183/195	255/267	295/295	235/244	278/278
ID3	BQ664	08/09/04	Pos	Pos	892	394	n/a	CM2	F	149/153	177/191	262/262	313/321	240/248	278/278

[1]INNO-LIA immunoblots are shown in Supplementary Fig. 2.
[2]Western blot profiles are shown in Fig. 1a.
[3]As shown in Supplementary Fig. 3 and Supplementary Table 3.
[4]F, female; as determined by amplification of a portion of the amelogenin gene (see methods).
[5]Microsatellite loci were amplified from fecal DNA; two alleles per locus are shown; homozygous loci were amplified a minimum of seven times to exclude allelic dropout.
[6]Partial degradation of sample CP465 precluded SIV amplification and complete microsatellite analysis; however, available data are consistent with CP465 representing the same individual as CP684 and CP685.
[7]n/a, not available.

Supplementary Table 3

Table S3. Mitochondrial DNA analysis of gorilla fecal samples.

Haplotype[1]	Fecal samples with identical mtDNA haplotype	GenBank accession number
CM1	BQ638	AM392403
CM2	BQ660, BQ662, BQ664, DP1153, DP1155, EK1178, KG1246	AM392404
CM3	BQ661	AM392405
CM4	CP465, CP684, CP685	AM392406
CM5	CP682, CP683, CP1438	AM392407
CM6	CP687, CP689	AM392408
CM7	CP1386	AM392409
CM8	CP1426	AM392410
CM9	CP1433, CP1434, CP1436, CP1437, CP1439, CP1447	AM392411
CM10	CP1435	AM392412
CM11	EK1104, KG1236	AM392413
CM12	EK1180	AM392414
CM13	EK1183	AM392415
CM14	KG1225	AM392416
CM15	BB71, BB228, BB242, NK743, NK1166	AM392417
CM16	LB18	AM392418
CM17	LB184	AM392419
CM18	LB212, LB712, LB794, LB1223	AM392420
CM19	LB700	AM392421
CM20	MB1207	AM392422
CM21	TK56	AM392423
CM22	TK62	AM392424
CM23	TK795	AM392425

[1] Haplotypes correspond to those shown in Supplementary Fig. 3. Haplotypes from SIVgor infected gorillas are highlighted in red.