Available online at www.sciencedirect.com
ScienceDirect
Virology 368 (2007) 155–171
VIROLOGY
www.elsevier.com/locate/yviro

Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon

Fran Van Heuverswyn [a], Yingying Li [b], Elizabeth Bailes [c], Cecile Neel [a,d], Benedicte Lafay [e], Brandon F. Keele [b], Katharina S. Shaw [b], Jun Takehisa [b], Matthias H. Kraus [b], Severin Loul [d], Christelle Butel [a], Florian Liegeois [a], Bienvenue Yangda [d], Paul M. Sharp [c], Eitel Mpoudi-Ngole [d], Eric Delaporte [a,f], Beatrice H. Hahn [b], Martine Peeters [a,*]

[a] *UMR145, Institut de Recherche pour le Développement (IRD) and Department of International Health, University of Montpellier 1, 911, Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France*
[b] *Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA*
[c] *Institute of Genetics, University of Nottingham, Nottingham, NH7 2UH, UK*
[d] *Projet Prevention du Sida au Cameroun (PRESICA), Yaoundé, Cameroun*
[e] *UMR CNRS-IRD 2724 (Génétique et Evolution des Maladies Infectieuses), Montpellier, France*
[f] *Department of Tropical and Infectious Diseases, University Hospital of Montpellier (CHU), Montpellier, France*

Received 7 May 2007; returned to author for revision 7 June 2007; accepted 13 June 2007
Available online 24 July 2007

Abstract

It is now well established that the clade of simian immunodeficiency viruses (SIVs) infecting west central African chimpanzees (*Pan troglodytes troglodytes*) and western gorillas (*Gorilla gorilla gorilla*) comprises the progenitors of human immunodeficiency virus type 1 (HIV-1). In this study, we have greatly expanded our previous molecular epidemiological survey of SIVcpz in wild chimpanzees in Cameroon. The new results confirm a wide but uneven distribution of SIVcpz*Ptt* in *P. t. troglodytes* throughout southern Cameroon and indicate the absence of SIVcpz infection in *Pan troglodytes vellerosus*. Analyzing 725 fecal samples from 15 field sites, we obtained partial nucleotide sequences from 16 new SIVcpz*Ptt* strains and determined full-length sequences for two of these. Phylogenetic analyses of these new viruses confirmed the previously reported phylogeographic clustering of SIVcpz*Ptt* lineages, with viruses related to the ancestors of HIV-1 groups M and N circulating exclusively in southeastern and south central *P. t. troglodytes* communities, respectively. Importantly, the SIVcpz*Ptt* strains from the southeastern corner of Cameroon represent a relatively isolated clade indicating a defined geographic origin of the chimpanzee precursor of HIV-1 group M. Since contacts between humans and apes continue, the possibility of ongoing transmissions of SIV from chimpanzees (or gorillas) to humans has to be considered. In this context, our finding of distinct SIVcpz*Ptt* envelope V3 sequence clades suggests that these peptides may be useful for the serological differentiation of SIVcpz*Ptt* and HIV-1 infections, and thus the diagnosis of new cross-species transmissions if they occurred.
© 2007 Elsevier Inc. All rights reserved.

Keywords: HIV; SIVcpz; Evolution; Prevalence; Genetic diversity; Cross-species transmission

Introduction

Numerous African primates are infected with simian immunodeficiency viruses (SIVs), and it is now well established that SIVs infecting chimpanzees (*Pan troglodytes troglodytes*) and western gorillas (*Gorilla gorilla gorilla*) in west central Africa are the progenitors of human immunodeficiency virus type 1 (HIV-1) (Bibollet-Ruche et al., 2004a; Gao et al., 1999; Keele et al., 2006; Van de Woude and Apetrei, 2006; Van Heuverswyn et al., 2006). The three groups of HIV-1 (groups M, N and O) are the result of three independent cross-species transmission events (Hahn et al., 2000; Sharp et al., 2005). Whereas infections with groups N and O remain mainly confined to Cameroon and surrounding countries, HIV-1 group M has spread worldwide and infected more than 60 million

* Corresponding author. Fax: +33 0467416146.
E-mail address: martine.peeters@mpl.ird.fr (M. Peeters).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.06.018

individuals (Ayouba et al., 2001; Hemelaar et al., 2006; McCutchan, 2006; Peeters et al., 1997; Vergne et al., 2003; Yamaguchi et al., 2006).

Chimpanzees (*Pan troglodytes*) are classified into four subspecies but only the two subspecies from central Africa, *P. t. troglodytes* and *Pan troglodytes schweinfurthii* are infected by SIVcpz. Moreover, their viruses form divergent subspecies-specific phylogenetic lineages, SIVcpz*Ptt* and SIVcpz*Pts*, respectively (Groves, 2001; Sharp et al., 2005). All HIV-1 strains fall within the SIVcpz*Ptt* lineage from west central Africa and no human counterpart has yet been identified for SIVcpz*Pts* in chimpanzees from east central Africa (Santiago et al., 2003a,b; Sharp et al., 2005; Worobey et al., 2004). Using non-invasive strategies to detect antibodies and viral RNA in ape fecal samples, we recently traced the natural reservoirs of HIV-1 groups M and N to distinct chimpanzee populations in southern Cameroon (Keele et al., 2006). In these communities, prevalence rates ranged from 0% to 35% and the new SIVcpz*Ptt* viruses exhibited a significant phylogeographic clustering, suggesting that major rivers or long distances were responsible for the uneven distribution of SIVcpz strains. Although, Cameroon is also home to the *Pan troglodytes vellerosus* chimpanzee subspecies, which inhabits the forests north of the Sanaga River, the boundary with the range of *P. t. troglodytes*, no case of SIVcpz infection has thus far been identified in this subspecies, but only a few have been sampled (Keele et al., 2006; Nerrienet et al., 2005).

Extension of our survey in Cameroon to a second great ape species, western gorillas (*Gorilla gorilla*), showed that these apes are also endemically infected with a simian immunodeficiency virus, designated SIVgor (Van Heuverswyn et al., 2006). Surprisingly, the phylogenetic relationships among HIV-1, SIVcpz and SIVgor indicate that the gorilla viruses form a monophyletic lineage within the SIVcpz*Ptt* radiation, which is much more closely related to HIV-1 group O than to any other SIV. Although not yet detected in chimpanzees, the SIVgor virus seemed to have a *P. t. troglodytes* origin, and it remains to be determined whether chimpanzees transmitted HIV-1 group O-like viruses to gorillas and humans independently, or first to gorillas which then transmitted the virus to humans.

In order to confirm and extend our previous findings and to study in more detail the genetic diversity of SIVcpz strains in wild chimpanzee populations, we conducted a comprehensive follow-up study to obtain more SIVcpz*Ptt* isolates from the same as well as from different geographic regions in Cameroon. We also characterized two new full-length genomes and analyzed V3 envelope sequences from additional SIVcpz*Ptt* strains. The latter provides new V3 peptides as potential diagnostic tools to discriminate between SIVcpz*Ptt* and HIV-1 infections.

Results

SIV infection of wild chimpanzees in Southern Cameroon

Fecal samples ($n=725$) were collected from wild-living (non-habituated) apes at 15 forest sites, primarily located in the southern part of Cameroon (Fig. 1). Collection sites, numbers of collected samples, species and subspecies determinations, and test results of SIVcpz antibody detection are summarized in

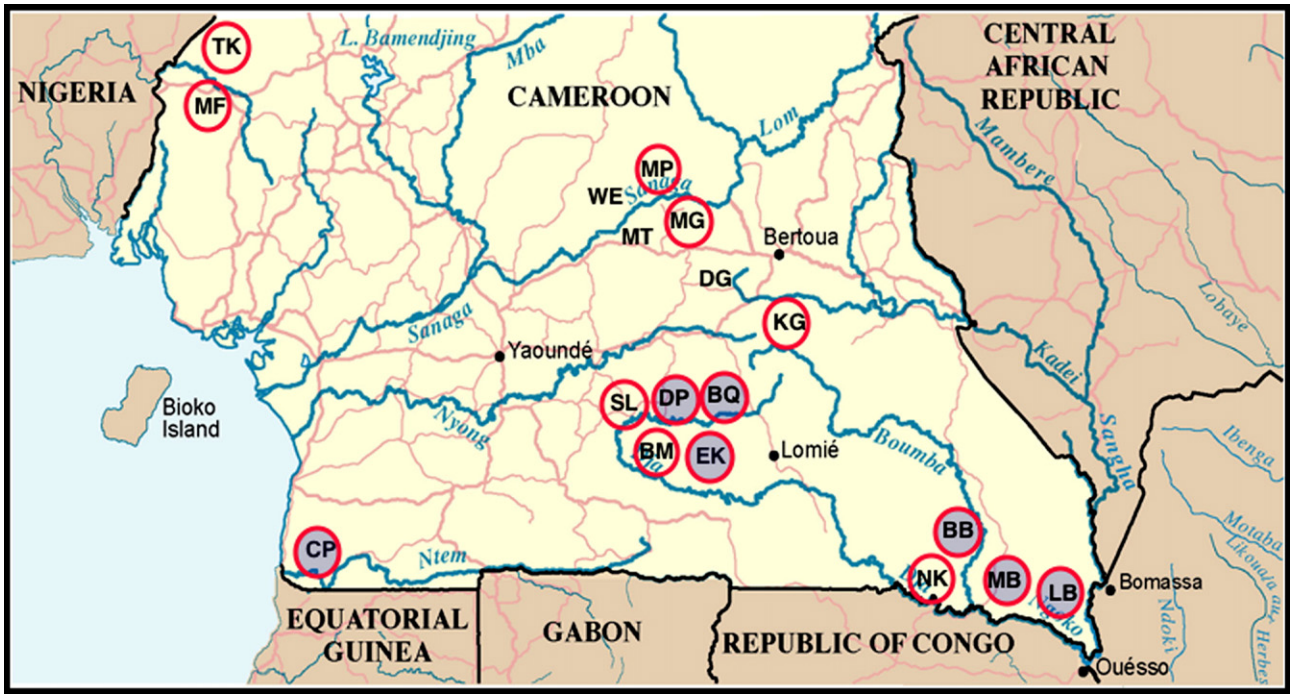

Fig. 1. Location of wild chimpanzee study sites in southern Cameroon. Circles indicate sampling sites from the current study ($n=15$). Circles with grey shading indicate field sites that were also studied previously (Keele et al., 2006). WE, MT and DG are sites included only in the last study (Keele et al., 2006). MF, TK, MP and WE are located in the range of *P.t.vellerosus*, while all other sites are located in the range of *P.t. troglodytes*.

Table 1. Of the 725 fecal samples, 378 were of chimpanzee origin based on mtDNA analysis, with 323 from *P. t. troglodytes*, collected south of the Sanaga River and 55 from *P. t. vellerosus*, all collected north of the river. The remaining specimens were either from gorillas (*G. gorilla*, $n=213$) and have been described in our previous study (Van Heuverswyn et al., 2006), or from other primate species ($n=75$) such as baboons, agile monkeys or different *Cercopithecus* species. For 59 samples, the DNA was degraded and the exact species of origin could not be determined. With the exception of BB (where only gorilla samples were collected), chimpanzee samples were obtained from all other sites.

All 378 chimpanzee samples were tested for the presence of SIVcpz antibodies with the INNO-LIA HIV I/II Score Confirmation Assay. We identified a total of 40 new SIVcpz antibody-positive samples, collected at 5 of the 11 sites located within the range of *P. t. troglodytes* apes (Table 1). These 40 specimens were also subjected to the enhanced chemiluminescent Western blot analysis and a similar reactivity profile was observed (Figs. 2a and b). All samples reacted strongly with the HIV-1 p24 core antigen, and some also did with p17. Samples from the LB and MB collection sites in the south east also exhibited strong cross-reactivity with HIV-1 envelope antigens; gp41 on INNO-LIA strips and gp41, gp120, gp160 on Western blot strips. The *env* cross-reactivity was weaker or absent for samples collected in the south central region (DP, SL, and BM). As observed previously, all Western blot profiles of chimpanzee samples from the southeastern corner of Cameroon were indistinguishable from the HIV-1-positive human plasma control.

In our previous study, samples were collected at 10 forest sites (WE, MT, DG, CP, DP, BQ, EK, BB, MB, and LB) and in 5 of them (MT, DP, EK, MB, and LB) SIVcpz*Ptt* infection was observed with prevalence rates ranging from 4% to 35% (Keele et al., 2006). For this survey, we obtained new chimpanzee samples from 6 of these 10 sites (CP, BQ, EK, DP, LB, and MB). In addition, we extended the sampling area to 5 new sites (MG, KG, SL, BM and NK) in the *P. t. troglodytes* range and 3 new sites (TK, MF, and MP) in the *P. t. vellerosus* range. Like the previously reported 23 *P. t. vellerosus* samples (Keele et al., 2006), none of the additional 55 samples collected north of the Sanaga River exhibited SIVcpz antibody reactivity, suggesting that this subspecies is not infected with SIVcpz.

We estimated the number of individuals that were sampled by taking into consideration the extent of sample degradation (0.169) and oversampling (on average, each chimpanzee was sampled 1.716 times) (Keele et al., 2006). From this, we calculated the prevalence of SIVcpz infection at the DP, SL, BM, LB, and MB sites. As observed before (Keele et al., 2006), the highest prevalence rates were seen in the southeast, 33.3% for MB and 34.5% for LB. In addition, moderate prevalence rates were observed at the other sites, 11.7% for DP, 5.4% for BM, and 4.7% for SL. In contrast to our previous study, no infected chimpanzees were found in EK, although we did identify infected chimpanzees at a new location, BM, within the Dja Reserve just below the Dja River and approximately 35 km west of EK. Furthermore, we identified positive samples at SL, located at the northern periphery of the Dja Reserve, just north of the Dja River, approximately 40 km west of DP. No evidence for SIVcpz infection was found in NK, similar to what was previously observed for BB, with both sites located west of the Boumba River. Finally, no positive samples were identified at KG and MG which were located in close proximity of DG and MT, respectively (Fig. 1), two sites where SIVcpz prevalence was low or absent in the previous study.

Microsatellite analyses revealed that the 40 immunoblot reactive samples represented 17 different *P. t. troglodytes* apes (Table 2). With the exception of 7 samples from MB, which were all derived from a chimpanzee (ID25) that had an identical

Table 1
SIV infection in wild chimpanzee populations from different field sites in Southern Cameroon

Collection sites [a]	Fecal samples collected	Chimpanzee samples [b]	*P. t. troglodytes* [b]	*P. t. vellerosus* [b]	Gorilla samples [b]	Other species [b]	Degraded samples [b]	SIVcpz antibody-positive samples	Number of SIVcpz-infected chimpanzees [c]
TK	25	1	0	1	21	1	2	0	0
MF	45	39	0	39	0	1	5	0	0
CP	107	42	42	0	51	10	4	0	0
MP	44	15	0	15	0	29	0	0	0
MG	25	24	24	0	0	1	0	0	0
KG	55	15	15	0	15	1	24	0	0
DP	70	35	35	0	16	7	12	10	2
BQ	35	10	10	0	25	0	0	0	0
SL	45	44	44	0	0	0	1	5	1
EK	43	27	27	0	6	9	1	0	0
BM	41	38	38	0	1	2	0	2	1
LB	70	18	18	0	48	2	2	3	3
MB	81	62	62	0	2	10	7	20	10
NK	34	8	8	0	25	1	0	0	0
BB	5	0	0	0	3	1	1	0	0
Total	725	378	323	55	213	75	59	40	17

[a] Location of sites is shown in Fig. 1.
[b] As determined by mitochondrial DNA analysis of fecal DNA (see Materials and methods).
[c] As determined by microsatellite analysis (Table 2).

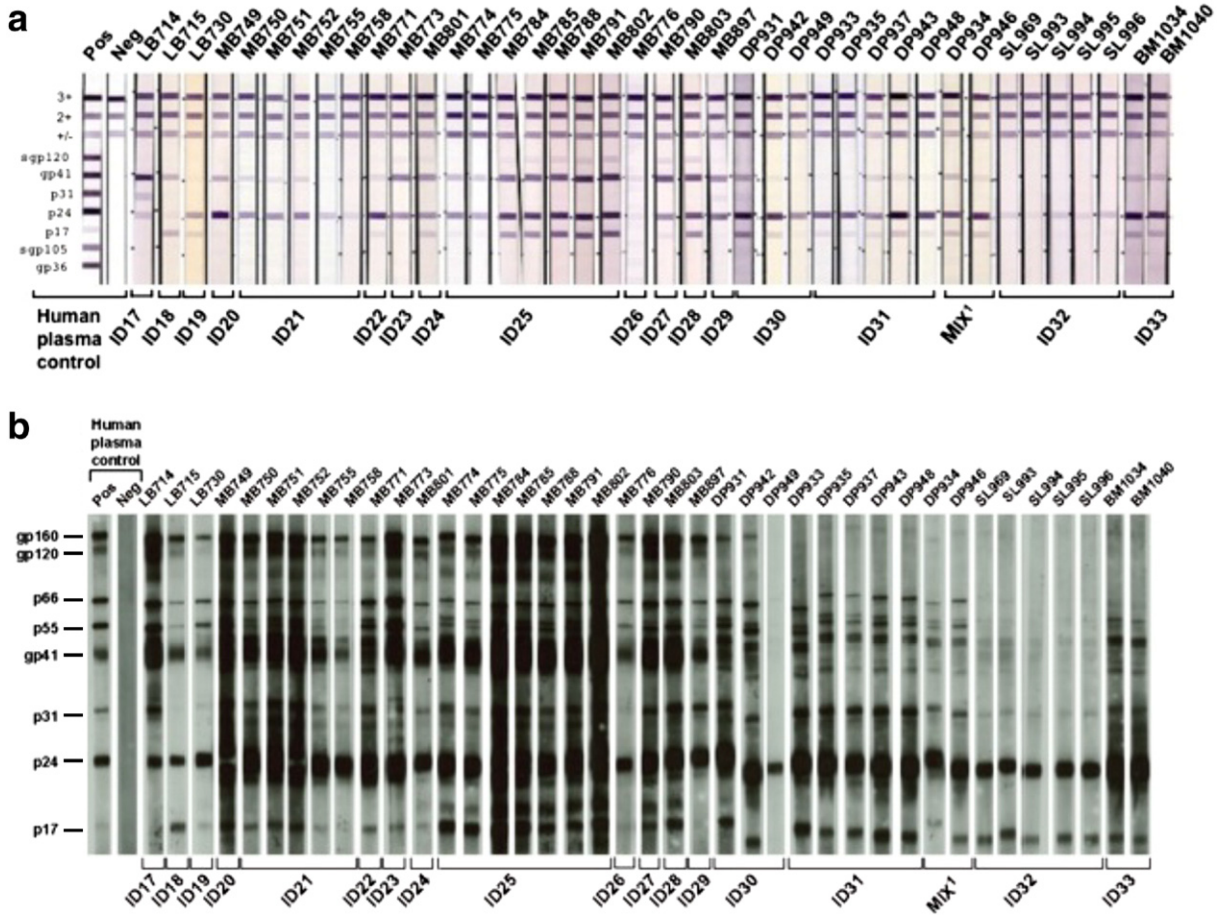

Fig. 2. Detection of HIV-1 cross-reactive antibodies in chimpanzee fecal samples. Fecal samples from wild-living chimpanzees were tested by the INNO-LIA HIV I/II Score Confirmation test (a) and ECL Western blot analysis using HIV-1 antigen containing strips (b). Samples are numbered with letters indicating their collection site as shown in Fig. 1. Samples from the same individual (ID) are grouped, individuals are numbered taking into account the previous survey (Keele et al., 2006) where 16 SIVcpz*Ptt* positive individuals were identified. Molecular weights of HIV-1 proteins are indicated. The banding pattern of plasma from an HIV-1 infected patient and an uninfected human control are shown.

microsatellite profile as a previously sampled individual (ID11) (Table 2), all other positive samples were derived from new individuals. For two antibody-positive samples, DP934 and DP946, individual assignment was not possible, because they represented a mixture of fecal material from individuals 14 and 15 from the same collection site.

Genetic diversity of SIVcpzPtt isolates from wild chimpanzee populations in partial pol and env regions

RNA was extracted and subjected to RT–PCR amplification using consensus *env* (gp41 and a fragment spanning the V2–V5 region) and *pol* primers. SIVcpz sequences were amplified for one or more genomic regions from 16 chimpanzees (Table 2), while despite strong antibody reactivity, one sample (MB773) was repeatedly virion RNA negative. SIVcpz sequences were obtained from all sites where antibody-positive chimpanzees were identified.

To compare the evolutionary relationships of the 16 new SIVcpz*Ptt* viruses to each other and to previously characterized SIVcpz, SIVgor and HIV-1 strains, phylogenies were constructed using partial Pol (296 aa) and Gp41 (126 aa) sequences (Figs. 3a and b). The relative positions of some strains, or clades, differed between the two trees, providing further evidence of recombination during the divergence of the SIVcpz*Ptt* lineages (Sharp et al., 2005). However, in both trees, all SIVcpz*Ptt* strains from MB and LB, including the 12 new strains, fell into a well-supported clade which included HIV-1 group M. In contrast, two new strains from DP clustered with strains obtained from captive chimpanzees in Cameroon (CAM3 and CAM5), rather than grouping closely with the two previously obtained strains from that same location. CAM3 is thought to have been caught south of the Dja River (Nerrienet et al., 2005), whereas the DP site lies north of the Dja River (Fig. 1); the precise location of capture of CAM5 is unknown. The two other strains were both from sites (SL and BM) not previously sampled. The new strain from SL was most closely related to strains previously obtained from MT, a site 130 km to the north of SL (Fig. 3c). The new strain from BM was most closely related to the clade composed of strains from EK and HIV-1 group N. Although we were not able to amplify *pol* sequences from this BM1034 sample, we obtained a 500 bp region from *gag*, which also clustered with the EK/HIV-1 group N clade (data not shown). The BM and EK sites both lie in the

Table 2
Genetic identification of SIVcpz infected chimpanzees

Individual[a]	Fecal samples	Date of sample collection	Fecal antibody detection[b]	Fecal vRNA detection[c]			MtDNA haplotype	Locus[d] D18S536	Locus[d] D4S243	Locus[d] D10S676	Locus[d] D9S922	Locus[d] D2S1326	Locus[d] D2S1333	Locus[d] D4S1627	Locus[d] D9S905
				pol	gp41	V3									
17	LB714	21/04/05	Pos	+	+	+	HT1	173/177	227/227	173/177	286/306	252/268	305/317	220/224	278/286
18	LB715	21/04/05	Pos	+	+	+	HT2	173/173	204/227	169/169	– / –	244/268	305/317	220/224	282/286
19	LB730	26/04/05	Pos	−	+	−	HT3	157/177	227/231	169/185	– / –	228/232	309/321	224/232	282/286
20	MB749	20/04/05	Pos	−	+	−	HT4	153/169	196/227	181/189	– / –	256/264	313/321	228/228	290/294
21	MB750	20/04/05	Pos	+	+	−	HT5	165/169	200/227	177/177	– / –	244/252	321/321	228/236	270/286
	MB751	20/04/05	Pos	+	+	−	HT5	165/169	200/227	177/177	– / –	244/252	321/321	228/236	270/286
	MB752	20/04/05	Pos	−	+	−	HT5	165/169	200/227	177/177	– / –	244/252	321/321	228/236	270/286
	MB755	20/04/05	Pos	−	+	−	ND	165/169	200/227	177/177	– / –	244/252	321/321	228/236	270/286
	MB758	20/04/05	Pos	−	−	−	ND	165/169	200/227	177/177	– / –	244/252	321/321	228/236	270/286
22	MB771	23/04/05	Pos	−	+	−	HT6	145/153	204/247	177/189	306/310	252/268	325/329	232/236	274/286
23	MB773	23/04/05	Pos	−	−	−	HT7	153/153	227/227	177/181	– / –	252/268	305/325	216/232	282/298
24	MB801	28/04/05	Pos	+	+	+	HT8	141/153	227/231	177/185	– / –	232/252	321/325	224/228	282/282
25[e]	MB774	23/04/05	Pos	−	−	−	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB775	23/04/05	Pos	−	−	−	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB784	24/04/05	Pos	−	−	−	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB785	24/04/05	Pos	−	−	−	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB788	28/04/05	Pos	+	−	+	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB791	28/04/05	Pos	−	−	−	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
	MB802	28/04/05	Pos	−	+	+	HT9	157/169	227/227	177/177	– / –	232/256	317/325	212/236	282/286
26	MB776	23/04/05	Pos	−	−	+	HT10	165/173	227/231	177/185	– / –	244/268	317/325	232/236	282/286
27	MB790	28/04/05	Pos	+	−	−	HT11	149/153	247/247	185/189	294/302	240/248	297/313	224/228	282/298
28	MB803	28/04/05	Pos	+	+	+	HT12	161/173	227/227	177/177	– / –	232/256	329/333	228/232	282/282
29	MB897	26/04/05	Pos	+	+	+	HT13	165/173	196/223	185/185	– / –	228/232	321/329	228/236	282/286
30	DP931	06/02/05	Pos	−	−	−	HT14	141/157	227/235	181/185	294/298	252/256	313/313	224/232	282/282
	DP942	06/02/05	Pos	−	+	−	HT14	141/157	227/235	181/185	294/298	252/256	313/313	224/232	282/282
	DP949	08/02/05	Pos	−	+	−	HT14	141/157	227/235	181/185	294/298	252/256	313/313	224/232	282/282
31	DP933	06/02/05	Pos	−	+	−	HT15	153/169	204/231	181/185	276/298	238/252	313/329	228/232	290/290
	DP935	06/02/05	Pos	+	−	+	HT15	153/169	204/231	181/185	276/298	238/252	313/329	228/232	290/290
	DP937	06/02/05	Pos	−	−	−	HT15	153/169	204/231	181/185	276/298	238/252	313/329	228/232	290/290
	DP943	06/02/05	Pos	+	+	−	HT15	153/169	204/231	181/185	276/298	238/252	313/329	228/232	290/290
	DP948	06/02/05	Pos	−	−	−	HT15	153/169	204/231	181/185	276/298	238/252	313/329	228/232	290/290
32	SL969	27/06/05	Pos	−	−	−	HT16	169/173	204/235	177/185	286/290	252/264	305/321	228/232	286/286
	SL993	28/06/05	Pos	−	−	−	HT16	169/173	204/235	177/185	286/290	252/264	305/321	228/232	286/286
	SL994	28/06/05	Pos	−	−	−	HT16	169/173	204/235	177/185	286/290	252/264	305/321	228/232	286/286
	SL995	28/06/05	Pos	−	+	+	HT16	169/173	204/235	177/185	286/290	252/264	305/321	228/232	286/286
	SL996	28/06/05	Pos	−	−	−	HT16	169/173	204/235	177/185	286/290	252/264	305/321	228/232	286/286
33	BM1034	21/06/05	Pos	−	+	+	HT17	137/161	248/231	177/181	– / –	240/244	313/321	224/224	286/290
	BM1040	21/06/05	Pos	−	+	−	HT17	137/161	248/231	177/181	– / –	240/244	313/321	224/224	286/290
Mix30/31	DP934	06/02/05	Pos	−	+	+	HT10	Mix	Mix	181/185	Mix	Mix	313/329	Mix	282/290
	DP946	06/02/05	Pos	+	−	−	ND	Mix	Mix	181/185	Mix	mix	313/329	Mix	282/290

ND: not done.

[a] Individuals are numbered in accordance to the previous survey (Keele et al., 2006) where 16 SIVcpz*Ptt*-positive apes were identified.

[b] Antibody profiles are shown in Fig. 2.

[c] *Pol* fragments of samples MB751, MB790 and DP935 are ~340 bp; the *pol* fragment of MB750 is 879 bp long; all other *pol* sequences are 888 bp in length.

[d] 8 STR loci were amplified from fecal DNA; 2 alleles per locus are shown; dashes indicate failure to amplify due to partial sample degradation; homozygous loci were amplified at least seven times to exclude allelic dropout.

[e] Note that the microsatellite profile of ID25 is identical to that of ID11 reported previously, thus indicating the resampling of the same individual.

Dja Reserve within the bend of the Dja River, only about 35 km apart (Fig. 3c).

With the addition of the new strains reported here, a new aspect of the phylogeographic diversity of SIVcpz*Ptt* is emerging. The MB and LB strains are all derived from a relatively small area spanning only about 100 km. For the partial Gp41 sequence, the average amino acid diversity among these strains is 15%, with a maximum pairwise difference of 26%. Although the DP strains do not form a monophyletic clade, their diversity is not very different from that of the MB/LB clade (average 16%, maximum 22%). Indeed, the gp41 clade composed of strains from DP, BM and EK, as well as previously characterized strains from captive chimpanzees from Cameroon (CAM3, CAM5, CAM13), Gabon (GAB1, GAB2) and an unknown origin (US), exhibits a similar level of diversity (average 18%, maximum 25%), although these strains were collected from a much wider geographic range. Thus, while there are, as expected, numerous examples of closely related strains at the same location reflecting local transmission, the results from the DP site indicate that divergent lineages may also be found at one location. The two lineages found at the DP site represent the diversity seen across a much wider area

including much of south central Cameroon, and even Gabon. In contrast, the strains present in southeastern Cameroon (the MB and LB sites) represent a much more isolated clade. Taken together, the new strains of SIVcpz*Ptt* reported here reinforce our earlier interpretation of the origins of HIV-1 groups M and N. In the gp41 tree, HIV-1 group M lies within the radiation of MB/LB strains and the phylogenetic separation of this clade from strains from other areas of west central Africa indicates that group M most likely originated in southeastern Cameroon. While HIV-1 group N lies within a larger clade of strains from widespread geographic origins, SIVcpz*Ptt* strains closely related to group N have only been found within the bend of the Dja River in south central Cameroon, pointing to that area as the likely source of that HIV-1 group.

Sequence analysis of two full-length SIVcpzPtt strains

To explore further the phylogeny of the SIVcpz*Ptt* lineage, we amplified two additional full-length SIVcpz*Ptt* strains from fecal samples collected at the MB (MB897) and DP (DP943) field

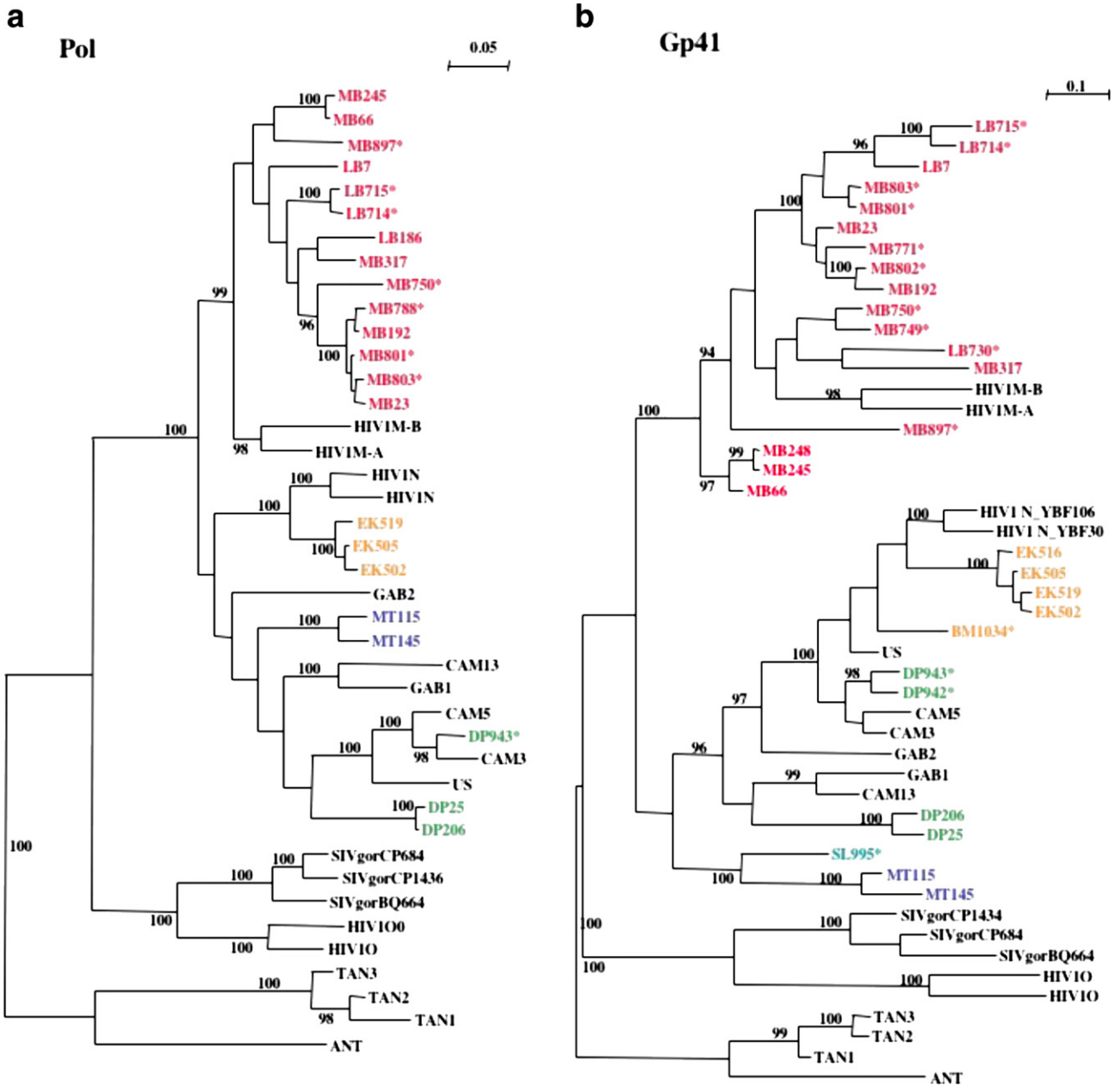

Fig. 3. Phylogenetic analysis of partial Pol (polymerase) (a) and Env (gp41, envelope transmembrane protein) (b) sequences of the newly identified SIVcpz*Ptt* strains (asterisks). Newly and previously identified SIVcpz*Ptt* strains (Keele et al., 2006) are highlighted by colors reflecting their collection sites (c). Representative HIV-1 group M (U455, LAI), N (YBF30, YBF106) and O (Ant70, MVP5180) sequences are included, as well as SIVcpz*Pts* sequences (ANT, TAN1-3) which form the outgroup. Trees were inferred by the Bayesian method; numbers on nodes are percentage posterior probabilities (only values above 95% are shown). The scale bars represent substitutions per site. Sites where positive samples were identified, as well as the exact location of these positive samples when GPS coordinates were available are shown (c). Locations where SIVcpz*Ptt* infection was found are color-coded to correspond with distinct SIVcpz*Ptt* lineages shown in (a and b). Strains from the previous survey are indicated in italics, while strains from the expanded survey are highlighted in bold letters.

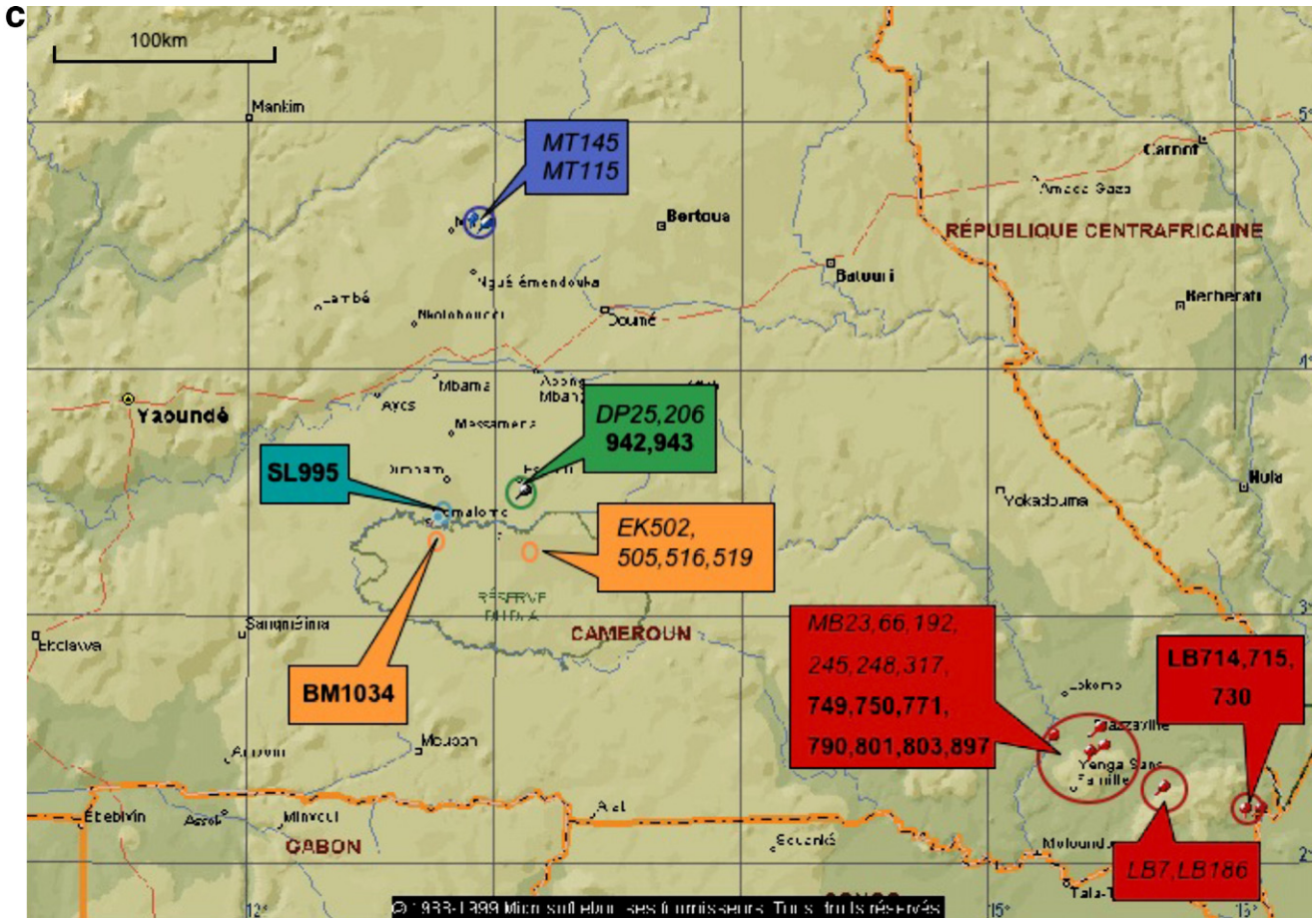

Fig. 3 (*continued*).

sites, respectively. Using previously described RT–PCR strategies (Santiago et al., 2003a,b; Keele et al., 2006; Takehisa et al., 2007), complete MB897 and DP943 genomes were amplified as sets of 11 and 14 partially overlapping subgenomic fragments, respectively (Fig. 4). Direct sequence analysis of these RT–PCR products yielded positions of base mixtures as well as nucleotide differences in adjoining fragments in regions of sequence overlap but none introduced nonsense or frameshift mutations (Table 3). Both MB897 and DP943 consensus sequences contained uninterrupted reading frames for all structural and regulatory proteins as well as intact regulatory elements.

Previous analyses showed that the phylogenetic relationships among SIVcpz*Ptt* strains vary in different regions of the genome (Keele et al., 2006), presumably reflecting recombination events during the divergence of the SIVcpz*Ptt* clade. Therefore we performed phylogenetic analyses on four regions of the proteome, namely Gag, the first part of Pol (sites 1–700), the remainder of Pol concatenated with Vif, and Env (Fig. 5c). Consistent with the analyses based on partial Pol and Env sequences, in all four trees MB897 formed a clade with the two other strains from the same geographic region (MB66 and LB7) and with HIV-1 group M, while DP943 formed a clade with three strains whose precise geographic origins are unknown (CAM3, CAM5 and US). However, within both clades, there were significant differences in branching order among the four trees. For example, MP987 clustered with MB66 in two trees (Pol1 and Env) but with LB7 in a third (Pol2), in each case with high probabilities. Similarly, DP943 clustered with CAM3 in two trees (Gag and Pol1), but outside a clade composed of CAM3 and CAM5 in the other two trees; these results were supported by high probability values in three of the four trees. These well-supported differences in branching order are evidence of comparatively recent recombination events during the divergence of viruses within these clades, and thus imply instances of dual infection.

There were also differences in branching order at deeper points within the SIVcpz*Ptt* clade (Fig. 5). For example, the clade including EK505 and HIV-1 group N grouped with the MB/LB/HIV-1 cluster in Gag and Pol1, but with another cluster (DP943/CAM3/CAM5/US) in Pol2 and Env, while the relative positions of MT145, GAB2 and the cluster of GAB1/CAM13 moved among all four trees. These results point to a number of recombination events during the earlier divergence of SIVcpz*Ptt* strains.

Genetic diversity among gp120 V3-loop sequences from different SIVcpz lineages

The V3 loop of the HIV-1 gp120 glycoprotein plays an important role in coreceptor binding of the virus, determining

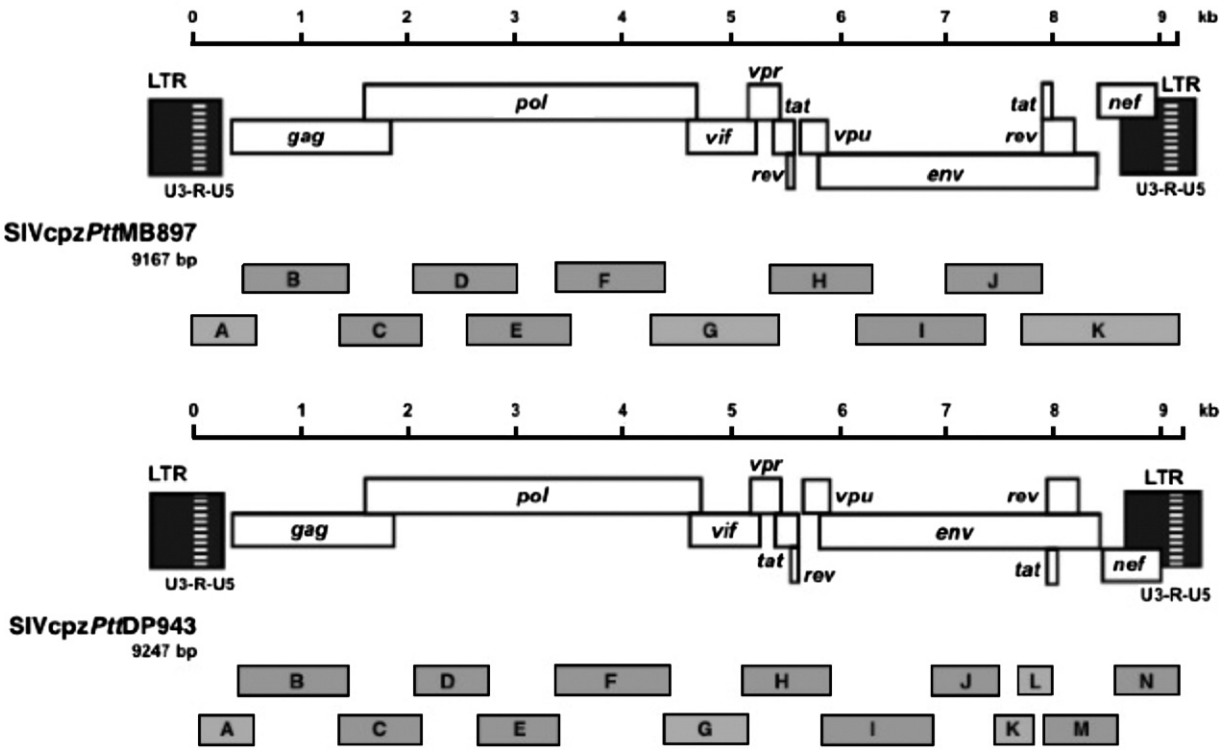

Fig. 4. Amplification of full-length SIVcpz*Ptt* (MB897 and DP943) genomes from fecal virion RNA. The complete MB897 and DP943 genomes were amplified as partially overlapping RT–PCR fragments (shaded boxes) from fecal virion RNA and sequenced without interim cloning (fragments are drawn to scale). Nucleotide sequences are numbered from the beginning of the R region in the 5′ LTR.

which coreceptor, CXCR4 or CCR5, is used for entry. Synthetic peptides derived from consensus V3-loop sequences of HIV-1 M, N and O can be used in ELISA assays to discriminate the different HIV-1 group infections (Ayouba et al., 2000, 2001; Mauclere et al., 1997; Simon et al., 2001; Vergne et al., 2003; Yamaguchi et al., 2004, 2006). The V3 loop of HIV-1 typically consists of 35 amino acids, but varies in length (range 31–39) and exhibits extensive sequence diversity, presumably in response to immune pressures. In contrast, V3 loop sequence diversity has been found to be relatively low among previously characterized SIVcpz sequences (Vanden Haesevelde et al., 1996; Hahn et al., 2000; Bibollet-Ruche et al., 2004b). In this study, we amplified the V2–V5 region for 10 new SIVcpz strains, extending the number of available SIVcpz V3 loop sequences to 27, with 23 derived from SIVcpz*Ptt* and 4 from SIVcpz*Pts*. Phenetic of these new sequences, as well as others reported in the past (Bibollet-Ruche et al., 2004b), reveals more diversity among SIVcpz V3 loop sequences than previously found (Fig. 6). With the exception of SIVcpz*Ptt*GAB2, which has an unusually divergent V3 loop, the SIVcpz sequences fall into two distinct phenetic clusters, one exclusively composed of SIVcpz*Ptt* strains from LB and MB, and the other containing all other SIVcpz strains, including SIVcpz*Pts* as well as HIV-1 group N sequences. These SIVcpz V3 sequences differ significantly from those of HIV-1 group M and O, and may thus represent distinct V3 serotypes. Among the SIVcpz*Ptt* variants most closely related to HIV-1 group M, the crown of the V3 loop is characterized by the GPGQ motif, which is the same in the human viruses. Interestingly, the second Gly found in the GPG motif of one of these SIVcpz strains (MB66) was replaced by a Ser; the biological significance of this change is unclear. Most other SIVcpz strains, including those from *P. t. schweinfurthii*, encode GPGM at the crown, while the HIV-1 group N strains and two SIVcpz*Ptt* strains (MT145 and DP935) encode GPAM.

Discussion

Our previous molecular epidemiological studies of wild-living apes in Cameroon allowed us to trace the origin of HIV-1 groups M and N to geographically isolated chimpanzee populations in southern Cameroon (Keele et al., 2006). We also identified for the first time SIV infection in wild gorillas and showed that this new lineage clustered within the SIVcpz radiation, as a sister clade to HIV-1 group O (Van Heuverswyn et al., 2006). In order to corroborate and extend these findings and to study in more detail the genetic diversity of SIVcpz, we conducted a more extensive survey of wild chimpanzee populations including previous and new field sites. The results from this new study confirm a wide but uneven distribution, as well as significant phylogeographic clustering, of SIVcpz*Ptt*. We confirmed the high prevalence (30%) of SIVcpz infection in southeastern Cameroon and the absence or low to moderate prevalence (4% to 11%) in chimpanzee communities from the south-central area. Finally, there is no evidence for SIVcpz infection in *P. t. vellerosus*.

Thus, a total of 824 chimpanzee samples have now been studied from wild chimpanzees in Cameroon, 746 from *P. t.*

Table 3
Positions of sequence ambiguities in MB897 and DP943 fecal population sequences

Virus	Amplicon [a]	Sequence ambiguity [b]	Position in genome	Protein [c]	Codon	Encoded amino acid
SIVcpzMB897	∗C	A/G	1766	p6 Gag	32	K/E
	∗C	A/G	1766	TF	47	E/G
	∗C	A/G	1783	p6 Gag	37	E
	∗C	A/G	1783	TF	53	T/A
	∗C–D	G/A–G	2076	PR	86	G
	∗E	A/C	2859	RT	248	E/D
	∗E	A/G	2995	RT	294	T/A
	H–I	C–T	6213	Env	140	P–S
	I–J	A–G	7116	Env	441	K–D
	I–J	A–T	7118	Env	441	K–D
	I–J	G–C	7124	Env	443	E–D
	I–J	A–C	7154	Env	453	E–D
	∗K	C/T	8108	Env	771	L
	∗K	C/T	8108	Rev	97	R/C
	∗K	C/T	8246	Env	817	I
SIVcpzDP943	∗B	A/G	687	p17 Gag	110	R
	∗B	C/T	1312	p24 Gag	190	L
	D–E	C–T	2711	RT	192	D
	E–F	A–G	3419	RT	428	Q
	∗I	C/G	6343	Env	174	Q,H,P/E,D,A
	∗I	A/C	6344	Env	174	Q,H,E,D/P,A
	∗I	G/T	6345	Env	174	Q,E,P,A/H,D,P,A
	∗I	A/C	6356	Env	178	N/T
	∗I	G/T	6360	Env	179	K/N
	∗I	A/G	6567	Env	248	Q
	∗I	C/T	6605	Env	261	T/I
	∗I	C/T	6608	Env	262	A/V
	∗I	C/G	6619	Env	266	H/D
	∗I	A/C	6626	Env	268	T/K
	∗I	A/G	6628	Env	269	K/E
	∗I	A/G	6637	Env	272	K/E
	∗I	A/G	6787	Env	322	I/V
	∗I	A/C	6816	Env	331	L/F
	∗I	A/T	6819	Env	332	K/N
	∗I	A/G	6826	Env	335	M/V
	K–L	A–G	7749	Env	642	Q

[a] The location of MB897 or DP943 amplicons is shown in Fig. 1.
[b] Dashes (–) indicate nucleotide differences between adjoining fragments in regions of sequence overlap; slashes (/) indicate positions of base mixtures; the latter are also highlighted by asterisks on the corresponding amplicon.
[c] p6, p17 and p24 Gag, viral core proteins; TF, Transframe peptide resulting from processing of Pr160 Gag–Pol polyprotein precursor; PR, protease; RT, reverse transcriptase; Env, envelope glycoprotein.

troglodytes and 78 from *P. t. vellerosus*, yielding a total of 31 new SIVcpz*Ptt* strains. All of these newly identified SIVcpz*Ptt* strains were found to fall within the radiation of SIVcpz from *P. t. troglodytes* apes, confirming that chimpanzees from east and west central Africa harbor subspecies-specific SIVcpz lineages.

Within Cameroon, chimpanzee populations that are separated by major geographical barriers also harbor distinct SIVcpz*Ptt* lineages. All SIVcpz*Ptt* strains from the MB and LB sites formed a single well-supported clade that included HIV-1 group M, and this clade was exclusively found in southeastern Cameroon. Similarly, samples from the EK and BM sites in the Dja Reserve, just south of the Dja River, were much more closely related to HIV-1 group N than were any other identified SIVcpz strains. While strains from neighbouring chimpanzee communities collected over a small area often clustered closely together in the phylogenies, it is also important to note that certain communities harbor rather divergent SIVcpz*Ptt* lineages. The most striking examples were LB714/715 and LB730, which were collected at a distance of 1 km from each other within the LB site, and DP25/DP206 and DP942/943 from DP, which were collected at a single field site. Moreover, the genetic diversity among SIVcpz*Ptt* strains from a relatively small region within the south east was comparable to that observed among all other SIVcpz*Ptt* strains, collected from sites across south-central Cameroon (DP, SL and MT), as well as from captive chimpanzees from the same area (CAM3, CAM5 and CAM13) and from Gabon (GAB1 and GAB2) (Corbet et al., 2000; Nerrienet et al., 2005; Peeters et al., 1989).

Our studies in Cameroon suggest that rivers may play a major role in the phylogeographic clustering of SIVcpz. For example, the SIVcpz*Ptt* strains most closely related to HIV-1 group N are found only in the north of the Dja Reserve, whereas just across the Dja River divergent lineages are found. Rivers could also explain the absence of SIVcpz infection in some areas. For example, the highest SIVcpz*Ptt* prevalences were observed at the MB and LB sites in south east Cameroon, while

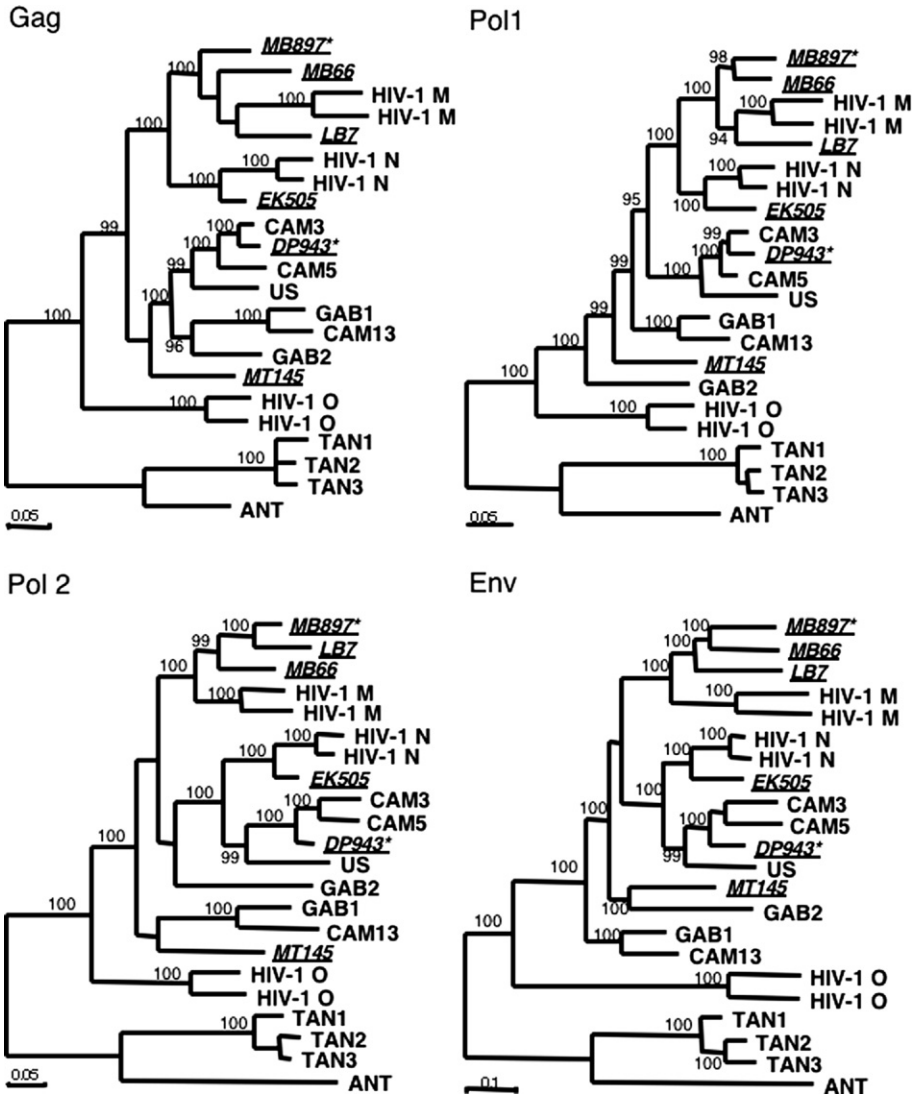

Fig. 5. Evolutionary relationships of full-length SIVcpz genomes. Phylogenetic analyses were performed on four major regions of the proteome: Gag (upper left), Pol sites 1–700 (Pol1, upper right), Pol sites 701–923 concatenated with Vif (Pol2, lower left), and Env (lower right). The new SIVcpz*Ptt* sequences are indicated by asterisk. Representative HIV-1 group M (U455, LAI), N (YBF30, YBF106) and O (Ant70, MVP5180) sequences are included, as well as SIVcpz*Pts* sequences (ANT, TAN1–3) which form the outgroup. Numbers on branches indicate estimated posterior probabilities (only values greater than 90% are shown). The scale bars represent 0.05 (or 0.1 for Env) amino acid replacements per site.

no sign of infection was detected at the BB and NK sites, separated from MB and LB by the Boumba River. Long distances may also play a role, although some viral lineages have clearly been able to spread over a wide range. Overall, and not unexpectedly, the genetic diversity among SIVcpz*Ptt* strains is heavily influenced by the geographic origin of their hosts.

We recently identified SIV in gorillas in Cameroon (Van Heuverswyn et al., 2006). These viruses (SIVgor) cluster within the SIVcpz*Ptt* lineage and are most closely related to HIV-1 group O. The phylogenetic position of SIVgor implies that *P. t. troglodytes* chimpanzees were the source of SIVgor. While we have now screened a substantial number of new samples from chimpanzees in Cameroon, including some from sites in areas which overlap with the range of the SIV-positive gorillas, we have not identified an SIVcpz*Ptt* strain which is closely related to SIVgor or HIV-1 group O. It is possible that, subsequent to transmission to gorillas, this lineage of SIVcpz*Ptt* has become extinct or had its origin outside Cameroon.

Two full-length genome sequences were characterized to verify the phylogenetic relationships among the different SIVcpz*Ptt* lineages. Analysis of these sequences confirmed previously reported ancient recombination events, for example involving the ancestors of the EK strains (and hence HIV-1 group N), as well as the GAB2, MT145 and GAB1/CAM13 strains (Keele et al., 2006; Sharp et al., 2005). These analyses also revealed evidence for more recent recombination events among the strains in southeastern Cameroon (at MB and LB). Thus, co-infection of individual chimpanzees by (at least) two strains of SIVcpz has occurred on multiple occasions.

Although highly endangered, gorillas and chimpanzees continue to be hunted, especially in west central Africa, and remain a potential source of human infection (Matthews and

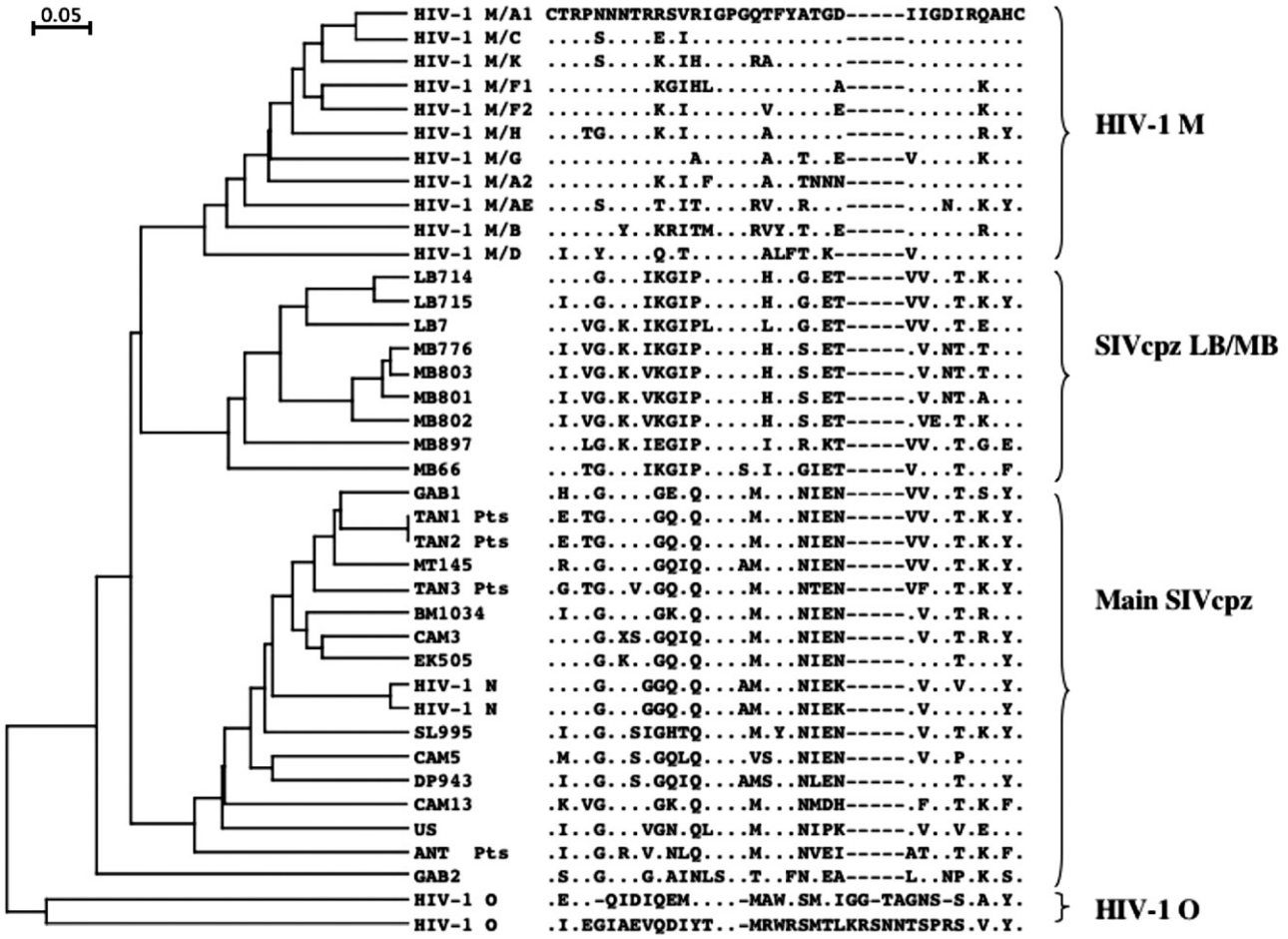

Fig. 6. Comparison of V3 loop sequences from the newly isolated SIVcpz with those of previously characterized representatives of the HIV-1/SIVcpz lineage. Single-letter abbreviations for amino acid residues are compared to HIV-1 M subtype A1 92UG037. Dots indicate amino acid identity and dashes indicate gaps introduced for better alignment. The phenetic similarities are indicated on the left.

Matthews, 2004). SIVs from chimpanzees (and perhaps gorillas) are now known to have crossed the species barrier to humans on at least three occasions (Hahn et al., 2000; Sharp et al., 2005). In addition, in West Africa there have been at least 8 cross-species transmissions of SIV from sooty mangabeys leading to distinct groups of HIV-2 (Damond et al., 2004; Santiago et al., 2005). Thus it would be not surprising if additional cases of SIVcpz or SIVgor cross-species transmissions had occurred, especially in geographic regions where these viruses are most prevalent and where hunting pressure is high. Since SIVcpz/SIVgor infections are unlikely to be differentiated by commercial HIV-1 antibody screening assays, such transmissions may have gone unrecognized. As shown in our serological testing, some SIVcpz*Ptt* variants elicit antibody responses that are indistinguishable from those observed in HIV-1 group M infections (see Western blot profiles of SIVcpz infections from the MB and LB sites in Fig. 2), or from the HIV indeterminate Western blot profiles which are often observed for HIV-1 group O infections (Gurtler et al., 1994; De Leys et al., 1990; Loussert-Ajaka et al., 1994). It is thus likely that some HIV-1 seropositive and/or seroindeterminant individuals in west central Africa are infected with the descendants of still other SIVcpz and/or SIVgor strains. Although only a fraction of human antibody-positive samples from west central Africa have been molecularly characterized, such studies have already allowed the identification and characterization of HIV-1 groups O and N in Cameroon (Gurtler et al., 1994; De Leys et al., 1990; Simon et al., 1998). V3 peptide ELISAs have been shown to reliably identify the majority of HIV-1 group M, N and O as well as dual M/O infections (Ayouba et al., 2000, 2001; Mauclere et al., 1997; Simon et al., 2001; Vergne et al., 2003; Yamaguchi et al., 2004, 2006). However, 10% of HIV-1-positive sera are untypeable (Vergne et al., 2003); while this is most probably related to the high specificity and the insufficient sensitivity of the assays, some of these samples may represent infections with other HIV-1 variants resulting from additional SIVcpz or SIVgor cross-species transmissions. It should be emphasized that HIV-1 group N was initially identified because the serum was non-reactive with HIV-1 M and O peptides, but recognized a peptide derived from SIVcpz (Simon et al., 1998). Subsequent genetic characterization of HIV-1 group N confirmed its close genetic relationship with SIVcpz*Ptt* in the envelope gene (Corbet et al., 2000; Roques et al., 2004; Simon et al., 1998). We identified a substantial number of new SIVcpz*Ptt* V3 loop sequences, which fall into distinct phenetic clusters and differ from those of HIV-1 groups M and O by 44% to 69%, and likely represent distinct V3 serotypes. This

degree of V3 loop diversity may allow the serological distinction of these SIVcpz*Ptt* lineages and identify possible new SIVcpz or SIVgor cross-species transmissions. If additional transmissions have indeed occurred, it will be important to study to what extent these new zoonoses have spread within the human population. Such events would have important public health implications and consequences for future AIDS treatment and vaccine development. In addition to virus, host and socio-demographic factors, high HIV-1 prevalences in certain rural areas, for example around logging concessions, could also play a role in the further spread of new transmissions, in particular by recombination between HIV-1 and SIVcpz or SIVgor (Laurent et al., 2004).

We have already sequenced full-length genomes for a subset of the new SIVcpz*Ptt* lineages. As a next step, it will be important to characterize these viruses on a biological level to elucidate the determinants of cross-species transmission and host adaptation. These analyses may also identify factors related to further spread of the virus; for example, it is currently unknown whether the global spread of HIV-1 group M, compared to the limited spread of HIV-1 groups N and O, reflects characteristics of the progenitors of these viral clades or simply different epidemiological opportunites that they encountered. While it seems that viral isolation from fecal material is not possible, we have recently demonstrated that infectious molecular clones of SIVcpz can be generated from fecal viral consensus sequences from wild-living chimpanzees (Takehisa et al., 2007). This new technology will allow the study of the biological characteristics of the new SIVcpzPtt strains, and yield better insights into their pathogenic potential.

In summary, the new SIVcpz*Ptt* strains reported in this study have corroborated and extended our previous findings regarding the origins of HIV-1 groups M and N in southeastern and south central Cameroon, respectively. We have observed a high genetic diversity, generated by both divergence and recombination, among SIVcpz*Ptt* strains from neighbouring chimpanzees communities, and confirmed the phylogeographic clustering of these viruses. Humans continue to have contact with the apes harboring these viruses, and the possibility of additional SIVcpz and/or SIVgor transmissions thus has to be considered.

Materials and methods

Sample collection and study sites

Fecal samples ($n=725$) were collected from wild-living (non-habituated) apes at 15 forest sites, mainly located in the southern part of Cameroon (Fig. 1). Twelve field sites were in the range of the *P. t. troglodytes* subspecies and three were north of the Sanaga River (TK, MF, and MP) in the *P. t. vellerosus* range (Table 1). Seven of the fifteen sites (EK, BM, CP, TK, BB, NK, and LB) were located in National Parks or Forest Reserves, while the remainder were in non-protected areas with considerable hunting pressure. In addition to the 7 previously sampled sites (CP, DP, BQ, EK, MB, BB, and LB) (Keele et al., 2006), 8 new field sites are included in this study: TK, MF, MP, MG, KG, SL, BM, and NK. Samples were collected by experienced trackers, preserved in RNA*later* (a commercial preservative of nucleic acids) (Ambion, Austin, TX), stored at the base camps at room temperature (maximum 20 days), and subsequently transported to a central laboratory in Yaounde. Collection site, date, global positioning system (GPS) coordinates (when available) and species origin, as identified by visual inspection were recorded in the field.

Species and subspecies determinations

The species origin of the fecal samples was determined by mitochondrial DNA (mtDNA) analysis as described previously (Keele et al., 2006; Van Heuverswyn et al., 2006). Fecal DNA was extracted using the QIAamp Stool DNA Mini kit (Qiagen, Valencia, CA). Briefly, 1.5 ml of fecal RNA*later* mixture was resuspended in stool lysis buffer and clarified by centrifugation. The supernatants were treated with an InhibitEx tablet (Qiagen, Valencia, CA), subjected to proteinase K digestion, and passed through a DNA binding column. Bound DNA was eluted in 50–150 µl elution buffer, and 5 µl (aliquots) were used for mitochondrial DNA amplification. 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′). Phylogenetic analysis of these D loop sequences allowed identification of all chimpanzee samples and their subspecies classification (*P. t. troglodytes* or *P. t. vellerosus*). Whereas the majority of gorilla samples could also be identified with this approach, some samples yielded amplification products of poor quality and were reanalyzed by amplifying a 386-bp mtDNA fragment spanning the 12SrRNA gene (using primers 12S-L1091 5′-AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT-3′ and 12S-H1478 5′-TGACTGCAGAGGGTGACGGGCGGTGTGT-3′). Sequence analysis of these fragments revealed that most of them were gorillas, but that some samples were not from apes and belonged to other non-human primates such as baboons or agile monkeys. Samples that scored negative in both amplification assays were considered as representing degraded material.

Microsatellite analyses

Fecal DNA was extracted from all SIVcpz antibody-positive samples for microsatellite analysis to determine the number of infected individuals as previously described (Keele et al., 2006). All samples were genotyped at 7 loci (D18s536, D4s243, D10s676, D2S1326, D2S1333, D4S1627 and D9S905)), with an additional locus (D9s922) amplified for a select number of samples. All PCR reactions were performed in duplicate. Individuals whose genotype appeared homozygous were amplified a minimum of seven times to exclude allelic drop out.

Detection of SIVcpz antibodies in RNAlater preserved fecal chimpanzee samples

All chimpanzee fecal samples were tested for the presence of HIV-1 cross-reactive antibodies, using the INNO-LIA HIV I/II

Score Confirmation test (Innogenetics, Ghent, Belgium) as previously described (Van Heuverswyn et al., 2006). This test configuration includes HIV-1 and HIV-2 recombinant proteins and synthetic peptides, coated as discrete lines on a nylon strip. INNO-LIA-positive samples were also tested by Western blot analysis as reported previously (Calypte Biomedical; Rockville, MD) (Keele et al., 2006). RNA*later* precipitated immunoglobulin is resolubilized by diluting fecal/RNA*later* mixtures (1.5 ml) with PBS–Tween 20 (7.5 ml), followed by inactivation of the mixture for 1 h at 60 °C, centrifugation (3500×*g* for 10 min) to clarify the solution, and then dialyzing it against PBS overnight at 4 °C. The reconstituted extracts were then subjected to immunoblot analysis. Importantly, all samples containing SIVcpz antibodies scored positive in both assays.

Amplification of SIVcpz sequences from fecal RNA

Fecal RNA was extracted from all samples with HIV cross-reactive antibodies and subjected to RT-PCR amplification using SIVcpz/HIV-1 consensus primers in *env* (gp41 ectodomain) (~390 bp) and *pol* (~340 bp or ~890 bp) regions as previously described (Keele et al., 2006; Van Heuverswyn et al., 2006). To amplify the V3 region, additional primer sets were designed, based on the phylogenetic clustering of the SIVcpz strains (all primers are listed in Table 4). cDNA was synthesized using the R1 primer, followed by nested PCR using primers F1/R1 and F2/R2. For one sample a small fragment (~520 bp) in *gag* was amplified with the following primer set: CPZ-GAGF1 (5′-ATGGGWGCGAGRGCGTC-3′), CPZ-GAGR1 (5′-GCTTCWGCTARNACYCTWGCCTTATG-3′), CPZ-GAGF2 (5′-ATGAAACATHTAGTWTGGGCMAG-3′) and CPZ-GAGR2 (5′-TCCCAHTCNGCDGCTTCYTCATTGAT-3′). All RT-PCR reactions were performed with the Expand Reverse Transcriptase and the Expand Long Template PCR system (Roche Diagnostics, Indianapolis, IN) according to the manufacturer's instructions. Briefly, 10 µl of fecal vRNA was used for cDNA synthesis and 20 U of RNase Inhibitor (Ambion, Austin, TX) was added to the RT–PCR mixture. The mixture was then incubated for 1 h at 42 °C, followed by 5 min at 95 °C to inactivate the enzyme. 10 µl of genomic DNA was used for first-round PCR amplifications and 5 µl of the first-round reaction was used for the nested PCR with second-round primers, F2/R2, by using the same thermocycling conditions. Mostly, PCR amplifications included 35 cycles of denaturation (94 °C, 20 s), annealing (50 °C, 30 s) and elongation (68 °C, 1 min) in a Peltier Thermal Cycler (PTC-200). For some amplifications PCR

Table 4
Primer sets used to amplify partial SIVcpz *pol* and *env* (gp41 and V3) sequences

Gene	Round of nested PCR	Primer	Size (b)	Sequence [a]	Reference [b]
pol	1	CPZ-pol-F1	823	CCAGCNCACAAAGGNATAGGAGG	a,b,c
		CPZ-pol-R1		ACBACYGCNCCTTCHCCTTTC	a,b,c,d
pol	2	CPZ-pol-F2	340	GGAAGTGGATACTTAGAAGCAGAAGT	b,c
		CPZ-pol-R2		CCAATYCCYCCYYTTYKYTTAAAATT	c
pol	1	CPZ-pol-F1a	1207	ACCTGGATNCCWGANTGGGA	c,d
		CPZ-pol-R1		ACBACYGCNCCTTCHCCTTTC	a,b,c,d
		or			
		CPZ-pol-F1b		GTTACCTGGGTACCTGAGTGGGA	d
		CPZ-pol-F1c	1228	TGGTGGDCWGANTAYTGGCA	d
		CPZ-pol-R1b		ACTGCHCCYTCWCCTTTCCACAG	d
pol	2	CPZ-pol-F2a	941	TWYTATGTWGATGGRGCAGC	c,d
		CPZ-pol-R2		CCAATYCCYCCYYTTYKYTTAAAATT	c
		or			
		CPZ-pol-F2b	1034	ACCTGGATHCCHGANTGGGA	d
env (gp41)	1	CPZ-gp41-F1	591	TCTTAGGAGCAGCAGGAAGCACTATGGG	d,e
		CPZ-gp41-R1		AACGACAAAGGTGAGTATCCCTGCCTAA	d,e
env (gp41)	2	CPZ-gp41-F2	467	ACAATTATTGTCTGGTATAGTGCAACAGCA	d,e
		CPZ-gp41-R2a		TTAAACCTATCAAGCCTCCTACTATCATTA	d,
		Or			
		CPZ-gp41-R2b	450	TCCTACTATCATTATGAATATTTTTATATA	d,e
env (V3)	1	M-V3-CPZ-F1	631	TTGAACCAATTCCYATWYAYTATTGTGC	
		M-V3-CPZ-R1		ATTSCTYTYCCTACYCTYTGCCA	
env (V3)	2	M-V3-CPZ-F2	603	TATTGTGCHCCAGCTGGWTTTGC	
		M-V3-CPZ-R2		TTACAATTTGTCTTATTCTGCAHGG	
env (V3)	1	N-V3-CPZ-F1	793	TTCCAATAYATTAYTGTGCACCACCAGG	
		N-V3-CPZ-R1		CCTGGTGCTACTCCTATGGGTTCTAT	
env (V3)	2	N-V3-CPZ-F2	511	CAATGYACMCATGGAATAAAVCCAGT	
		N-V3-CPZ-R2		GCATAAATTCCTTTTCCTACYCTTGTCCA	
env (V3)	1	DP-V3-CPZ-F1	970	TCCTTTGAGCCAATTCCAATACA	
		DP-V3-CPZ-R1		CTGCGCCCATAGTGCTTCCTGCTGC	
env (V3)	2	DP-V3-CPZ-F2	919	CCAATTCCAATACAYTAYTGTGCA	
		DP-V3-CPZ-R2		AACAGYGCHCCTAGTCCAAAGGCT	

[a] Y=C/T, W=A/T, R=A/G, H=A/C/T, B=C/G/T, S=G/C, K=G/T, D=A/G/T, N=A/C/T/G.
[b] a: Courgnaud et al., 2001; b: Santiago et al., 2003b; c : Van Heuverswyn et al., 2006; d: Keele et al., 2006; e: Yang et al., 2000.

conditions were slightly modified (annealing temperatures <50 °C and/or touch-down PCR strategy). Extension times varied depending on the size of the expected fragment and were typically set at 1 min/kb. The resulting amplification products were gel purified (Qiagen, Valencia, CA) and directly sequenced using an automated sequencer (3130xl Genetic Analyser, Applied Biosystems, Foster City, CA). The partial *pol*, gp41 and V3-loop sequences of the new SIVcpz*Ptt* strains are available at GenBank under accession numbers: AM696210–AM696248.

Generation of full-length fecal SIVcpzPtt population sequences

SIVcpz*Ptt* MB897 and DP943 genomes were amplified from SIVcpz antibody-positive fecal samples collected at the MB (sampled 26/04/2005) and DP (sampled 06/02/2005) field sites,

Table 5
Oligonucleotide primers used to amplify full-length MB897 and DP943 viral sequences

Fragment[a]		Forward primer[b]		Reverse primer	Amplicon[c]
MB897					
A	CON-A-F1	ACTGGGTCTCTCTKGTTAGACC	MB897-A-R1	CTTTTACTCTAATTCTTTGATG	
	CON-A-F2	GTCTCTCTKGTTAGACCAGATT	MB897-A-R2	TTCTGATCCTGTTGTGAGAGCTGG	553
B[d]	CON-B-F1	ATGGGTGCGAGAGCGTC	CON-B-R1	TCYTTKCCACAATTRAARCA	
	CON-B-F2	ATGAAACATTTAGTATGGGCAAG	CON-B-R2	GCTTCWGCTARNACYCTWGCCTTATG	985
C[d]	CON-C-F1	GAWGTRAARAMYTGGATGAC	CON-C-R1	ATYTTYCCTTCYTKYTCCAT	
	CON-C-F2	CARAATGCNAAYCCAGA	CON-C-R2	CTGGYTTYAATTTKACTGG	855
D	MB897-D-F1	GGGAATTGGAGGTTTTATAAAAGTAAGAC	MB897-D-R1	CAGATGACTATACTTTCCAGGGCTACT	
	MB897-D-F2	AAAGAGCAATAGGTACAGTATTAGTAGGGC	MB897-D-R2	TAACTCTAATTCTGCTTCCTGTGTGAG	998
E	MB897-E-F1	TCCAGCTAGGAATACCACACCC	MB897-E-R1	TGAATGATCCCTAATGCATATTGTG	
	MB897-E-F2	CCAGTACATGGATGATCTATATGTAGGATC	MB897-E-R2	TCCCTCTATCAGTTACATATCCTGCTTT	843
F[d]	CPZ-pol-F1C[e]	TGGTGGDCWGANTAYTGGCA	CPZ-pol-R1[e]	ACBACYGCNCCTTCHCCTTTC	
	CPZ-pol-F2b[e]	ACCTGGATHCCHGANTGGGA	CPZ-pol-R2[e]	CCAATYCCYCCYYTTYKYTTAAAATT	1009
G	MB897-G-F1	GAATTTGGAATTCCCTACAATCC	CON-G-R1	AGTTTTAGGCTGACTTCCTGGATG	
	MB897-G-F2	CAATGAACAAAGAGTTAAAGAAAAT	CON-G-R2	TCTARRYTAGGATCTAYTGGCTCCAT	1169
H	CON-H-F1	GCTATMATAAGAATYCTGCAACAACT	MB897-H-R1	TTTGGACAAGCCTGGGTTATGGCTG	
	CON-H-F2	TGYCAWCATAGCAGAATAGGCAT	MB897-H-R2	CTTCTTATCTCTTAATTCTGT	919
I	MB897-I-F1	GCATGGAACAATAACATGGTAGACCAAATG	MB897-I-R1	TTCTGGCCTGTACCGTCAGCGT	
	MB897-I-F2	GGGATCAAAGCCTAAAGCCATGTG	MB897-I-R2	GCGCCCATAGTGCTTCCTGCT	1193
J	MB897-J-F1	GTGCAGAATAAGACAAATTGTAA	MB897-J-R1	AAGGGGTCTGGAACGACAAAGGTG	
	CON-J-F2	TGGCAAAGAGTAGGGAAAGGAAT	CPZ-gp41-R2b[e]	TCCTACTATCATTATGAATATTTTTATATA	789
K	MB897-K-F1	GAAAGAGATTGATAACTACACAGA	CON-K-R1	AAGGCAAGCTTTATTGAGGC	
	MB897-K-F2	TGGTCAAGCCTGTGGAATTGGTT	CON-K-R2	CACTCAAGGCAAGCTTTAT	1428
DP943					
A	DP943-A-F1	GGTCTCTCTTGCTAGACCAG	CON-A-R1	CCTTCTGATCCTGTTKTGAGAGCTGG	
	DP943-A-F2	GACCAGATTAGAGCCCGGGA	DP943-A-R2	GTTCTAACTGTTTTAATAGCTG	540
D	DP943-D-F1	AGAAGCTCTGCTAGACACAGGAGC	CON-D-R1	GRAGCTCATAGCCCATCCACAA	
	DP943-D-F2	GGCACAGTGTTGGTGGGACC	CON-D-R2	CAGGRGTAGTAAAMCCCCAAGT	733
E	CON-E-F1	CAAGGATGGAAAGGGTCACCRGCAAT	DP943-E-R1	CTCTGTTGGCTGCCCCATCTACA	
	CON-E-F2	CCATTCAGAMARCAGCATCCAGA	DP943-E-R2	CTGCCCCTGGTATGGGTTCTG	805
G	CON-G-F1	AGTGCTGCAGTTAAGGCAGCCTGTTGGTGG	CON-G-R1	CCTGGRTGNWKCCAGGGCTC	
	CON-G-F2	ATGGCAGTATTCATTCACAATTTT	CON-G-R2	CTTGTTCCATCTATCTTCT	808
H	DP943-H-F1	GCCATCAGAAAAGCTGTTTTAGGG	DP943-H-R1	ACACAGGCTTGTGAGGCCCAAAT	
	DP943-H-F2	GGGCACCAACAGGTAGGGTCCC	CON-H-R2	AGTGTTGTKTCTGCCTCTTTCCA	920
I	DP943-I-F1	GAGCAAGACAGGACGATAGTGGAA	DP943-I-R1	AATTCCCTTGCCCACTCTCATCCAGG	
	DP943-I-F2	TGATGGAGAGGATCAGAGGCAGC	DP943-I-R2	GAACAATGGGGTGGTGTTAC	1130
J	DP943-J-F1	GCATATTGTAATGTCAGCGCCACAG	DP943-J-R1	CCTGAGCAGCCCCACAGACCTAGG	
	DP943-J-F2	CACAGATCAACAAGGAGGAGACCCGG	DP943-J-R2	CTGTTTAATGCCCCATACTGAGAGTTGC	643
K	CPZ-gp41-F1[e]	TCTTAGGAGCAGCAGGAAGCACTATGGG	CPZ-gp41-R1[e]	AACGACAAAGGTGAGTATCCCTGCCTAA	
	CPZ-gp41-F2[e]	ACAATTATTGTCTGGTATAGTGCAACAGCA	CPZ-gp41-R2b[e]	TCCTACTATCATTATGAATATTTTTATATA	448
L	DP943-L-F1	CCTAGGTCTGTGGGGCTGCTCAGG	DP943-L-R1	GGAGGTTCCGGAGATCGTCCC	
	DP943-L-F2	CCTAACATGGCAAGACTGGGAC	DP943-L-R2	GGATGTGCCTCTGCCTGGCTCTCCACC	361
M	CON-M-F1	AAATGGCTGTGGTATATAAAAAT	CON-M-R1	TARCCCWTCCAGTCCYCCCTT	
	CON-M-F2	GCTTAAGAAAGGTTAGGCAGGG	CON-M-R2	CCYCCCTTWYYTYTTAAAAA	771
N	DP943-N-F1	CAAACTCTAGCATGGCTAGAAG	CON-N-R1	CAAGGCAAGCTTTATTGAGGC	
	DP943-N-F2	GGCTTTCCAGTTAGACCACAAGTACC	CON-N-R2	TTGAGGCTTAAGCAGTGGGTTC	637

[a] See Fig. 4 for position of individual fragments in the MB897 and D943 genomes.
[b] CON primers were designed according to HIV-1/SIVcpz consensus sequences. MB897 and DP943 primers are strain specific. F1, first-round forward primer; F2, second-round forward primer; R1, first-round reverse primer; R2, second-round reverse primer. R=A/G, Y=C/T, M=A/C, K=G/T, W=A/T, H=A/C/T, B=C/G/T, D=A/G/T, and N=A/C/T/G.
[c] Size of PCR amplicons.
[d] Fragments B, C and F of MB897 and DP943 were amplified using identical primer sets.
[e] For primer designation, also see Table 4.

essentially as described (Santiago et al., 2003a,b; Keele et al., 2006; Takehisa et al., 2007). Briefly, cDNA was synthesized by adding fecal vRNA (10 μl) to a RT-PCR master mix containing 1× Buffer (Invitrogen; Carlsbad, CA), 0.5 mM dNTP, 5 mM dithiothreitol, 2 pmol of primer, 20 U of RNase inhibitor (Ambion, Austin, TX), and 200 U of SuperScript RT III (Invitrogen), and by incubating the mixture for 3 h at 50 °C. MB897 cDNA (6 μl) was added to a PCR mix consisting of 1× Expand Buffer II (Roche Diagnostics; Indianapolis, IN), 0.5 mM dNTP, 500 nM of PCR primers, 25 μg of bovine serum albumin and 3.75 U of Expand High-Fidelity polymerase mixture (Roche Diagnostics). Amplification conditions for both first-round and nested PCR included 35 cycles of denaturation (94 °C, 0.5 min), annealing (53–60 °C, 0.5 min), and extension (68 °C, 1 min). DP943 amplification was conducted similarly except for using 10 μl cDNA, 0.4 mM dNTP, 400 nM primers and amplification conditions that included 55 cycles of denaturation (94 °C, 0.5 min), annealing (45–52 °C, 0.5 min) and extension (68 °C, 1.5 min). The primers used are listed in Table 5. Amplicons were gel-purified and sequenced directly using an ABI 3730 DNA Analyzer using Sequencer version 4.6 (Gene Codes Corporation, Ann Arbor, MI). Chromatograms were carefully examined for positions of base mixtures (Table 3). In addition, nucleotide differences between adjoining fragments in regions of sequence overlap were recorded (Table 3). The full-length MB897 (9167 bp) and DP943 (9247 bp) population sequences are available at GenBank under the accession numbers EF535994 and EF535993, respectively.

Phylogenetic analyses of SIVcpz sequences

The new SIVcpz*Ptt* nucleotide sequences were translated and compared to various previously published sequences (see below). Protein sequences were aligned using ClustalW (Thompson et al., 1994); where necessary, minor manual adjustments were performed in SeaView (Galtier et al., 1996) and sites that could not be unambiguously aligned or contained a gap in any sequence were excluded from the analyses. Phylogenies were inferred by the Bayesian method (Yang and Rannala, 1997), implemented in MrBayes version 3.1 (Ronquist and Huelsenbeck, 2003), run for 1,000,000 generations. Parameters were examined with the Tracer program (http://evolve.zoo.ox.ac.uk/software.html/id%3Dtracer). For the analyses of partial Pol and gp41 sequences, the numbers of amino acid sites examined were 296 and 126, respectively. Using the mixed model in MrBayes indicated that the rtREV model of amino acid change (Dimmic et al., 2002) was most appropriate; this model was used with gamma-distributed rates across sites. For the analyses of full-length sequences, four major regions of the proteome were analyzed: Gag, Pol1, Pol2, and Env. The Pol sequence was divided at the position of a recombination breakpoint previously identified in HIV-1 group N (Gao et al., 1999). In Pol2, the carboxy-terminal region of Pol was concatenated with Vif; in the region of overlap between Pol and Vif, only the Vif sequence was included. The JTT model of amino acid change (Jones et al., 1992) was used, with gamma distributed rates across sites. For the analysis of the V3-loop sequences, a phenetic dendrogram was estimated by the UPGMA method from a distance matrix of uncorrected sequence differences using NEIGHBOR from the PHYLIP package version 3.64 (Felsenstein, 2004).

GenBank accession numbers for full-length sequences used in comparative analyses are as follows: HIV-1 group M subtype A U455 (M62320) and subtype B HXB2 (K03455), HIV-1 group N YBF30 (AJ006022) and YBF106 (AJ271370) HIV-1 group O: ANT70 (L20587) and MVP5180 (L20571); SIVcpz*Ptt* CAM3 (AF115393), CAM5 (AJ271369), CAM13 (AY169968), EK505 (DQ373065), GAB1 (X52154), GAB2 (AF382828), LB7 (DQ373064), MB66 (DQ373063), MT145 (DQ373066), US (AF103818); SIVcpz*Pts*: ANT (U42720), TAN1 (AF447763), TAN2 (DQ374657), TAN3 (DQ374658). GenBank accession numbers for additional partial *pol* and gp41 sequences used in comparative analyses are as follows: SIVcpz*Ptt* MT115 (DQ370395, DQ370370), DP206 (DQ370403, DQ370375), DP25 (DQ370405, DQ370378), EK502 (DQ370408, DQ370381), EK516 (DQ370382), EK519 (DQ370411, DQ370383), MB23 (DQ370413, DQ370388), MB192 (DQ370415, DQ370392), MB317 (DQ370416, DQ370387), MB245 (DQ370417, DQ370385), LB186 (DQ370418), MB248 (DQ370386); SIVgor: BQ664 (AM296488, AM296484), CP684 (AM296489, AM296485), CP1434 (AM296487), CP1436 (AM296491). GenBank accession numbers for additional V3 loop sequences used in comparative analyses are as follows: HIV-1 group M: A1 92UG037 (U51190), A2 97CDKTB48 (AF286238), B BK132 (AY173951), C ETH2220 (U46016), D 94UG114 (U88824), 01-AE CM240 (U54771), F1 VI850 (AF077336), F2 02CM_0016BBY (AY371158), G DRCBL (AF084936), H VI991 (AF190127), K EQTB11C (AJ249235).

Acknowledgments

We thank the Cameroonian Ministries of Health, Environment and Forestry, and Research for permission to perform this study; the staff from the PRESICA project for logistical support; Leonard Usongo, Desire Dontego, Fouda Espedi, and Bertin Tshikangwa (World Wildlife Fund), Gilles Etoga and Donjouma M'bohand (Ministry of Environment and Forestry) and the staff from Projet Grands Singes for assistance in the field. This work was supported in part by the National Institutes of Health (R01 AI 50529, R01 AI 58715, P30 AI 27767), the Bristol Myers Freedom to Discover Program, the Agence National de Recherches sur le SIDA, France (ANRS, ANRS 12125) and the Institut de Recherche pour le Développement (IRD).

References

Ayouba, A., Souquieres, S., Njinku, B., Martin, P.M., Muller-Trutwin, M.C., Roques, P., Barre-Sinoussi, F., Mauclere, P., Simon, F., Nerrienet, E., 2000. HIV-1 group N among HIV-1-seropositive individuals in Cameroon. AIDS 14, 2623–2625.

Ayouba, A., Mauclere, P., Martin, P.M., Cunin, P., Mfoupouendoun, J., Njinku, B., Souquieres, S., Simon, F., 2001. HIV-1 group O infection in Cameroon, 1986 to 1998. Emerg. Infect. Dis. 7, 466–467.

Bibollet-Ruche, F., Bailes, E., Gao, F., Pourrut, X., Barlow, L., Clewley, J.P., Mwenda, J., Langat, D.K., Chege, G.K., McClure, H., Mpoudi-Ngole, E.,

Delaporte, E., Peeters, M., Shaw, G.M., Hahn, B., 2004a. New simian immunodeficiency virus infecting De Brazza's monkeys (*Cercopithecus neglectus*): Evidence for a Cercopithecus monkey virus clade. J. Virol. 78, 7748–7762.

Bibollet-Ruche, F., Gao, F., Bailes, E., Saragosti, S., Delaporte, E., Peeters, M., Shaw, G.M., Hahn, B.H., Sharp, P.M., 2004b. Complete genome analysis of one of the earliest SIVcpz*Ptt* strains from Gabon (SIVcpzGAB2). AIDS Res. Hum. Retroviruses 20, 1377–1381.

Corbet, S., Muller-Trutwin, M.C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F., Mauclere, P., 2000. *env* sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. J. Virol. 74, 529–534.

Courgnaud, V., Pourrut, X., Bibollet-Ruche, F., Mpoudi-Ngole, E., Bourgeois, A., Delaporte, E., Peeters, M., 2001. Characterization of a novel simian immunodeficiency virus from Guereza Colobus (*Colobus guereza*) in Cameroon: a new lineage in the nonhuman primate lentivirus family. J. Virol. 75, 857–866.

Damond, F., Worobey, M., Campa, P., Farfara, I., Colin, G., Matheron, S., Brun-Vezinet, F., Robertson, D.L., Simon, F., 2004. Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. AIDS Res. Hum. Retroviruses 20, 666–672.

De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wouters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B., Taelman, H., van der Groen, G., Piot, P., Tersmette, T., Huisman, J.G., Van Heuverswyn, H., 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. J. Virol. 64, 1207–1216.

Dimmic, M.W., Rest, J.S., Mindell, D.P., Goldstein, R.A., 2002. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. J. Mol. Evol. 55, 65–73.

Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle. Distributed by the author.

Galtier, N., Gouy, M., Gautier, C., 1996. SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548.

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

Groves, CP., 2001. Primate taxonomy. Smithsonian Series in Comparative Evolutionary Biology. Smithsonian Institution Press, Washington and London.

Gurtler, L.G., Hauser, P.H., Eberle, J., von Brunn, A., Knapp, S., Zekeng, L., Tsague, J.M., Kaptue, L., 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J. Virol. 68, 1285–1581.

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

Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20, W13–W23.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices form protein sequences. Comput. Appl. Biosci. 8, 275–282.

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

Laurent, C., Bourgeois, A., Mpoudi-Ngole, E., Butel, C., Peeters, M., Delaporte, E., 2004. Commercial logging and HIV epidemic, rural Equatorial Africa. Emerg. Infect. Dis. 10, 1953–1956.

Loussert-Ajaka, I., Ly, T.D., Chaix, M.L., Ingrand, D., Saragosti, S., Courouce, A.M., Brun-Vezinet, F., Simon, F., 1994. HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. Lancet 343, 1393–1394.

Matthews, A., Matthews, A., 2004. Survey of gorillas (*Gorilla gorilla gorilla*) and chimpanzees (*Pan troglodytes troglodytes*) in Southwestern Cameroon. Primates 45, 15–24.

Mauclere, P., Damond, F., Apetrei, C., Loussert-Ajaka, I., Souquiere, S., Buzelay, L., Dalbon, P., Jolivet, M., Mony-Lobe, M., Brun-Vezinet, F., Simon, F., Barin, F., 1997. Synthetic peptide ELISAs for detection of and discrimination between group M and group O HIV type 1 infection. AIDS Res. Hum. Retroviruses 13, 987–993.

McCutchan, F.E., 2006. Global epidemiology of HIV. J. Med. Virol. 78, S7–S12.

Nerrienet, E., Santiago, M.L., Foupouapouognigni, Y., Bailes, E., Mundy, N.I., Njinku, B., Kfutwah, A., Muller-Trutwin, M.C., Barre-Sinoussi, F., Shaw, G.M., Sharp, P.M., Hahn, B.H., Ayouba, A., 2005. Simian immunodeficiency virus infection in wild-caught chimpanzees from Cameroon. J. Virol. 79, 1312–1319.

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

Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J.P., Badombena, W., Luo, N., Vanden Haesevelde, M., Delaporte, E., 1997. Geographical distribution of HIV-1 group O viruses in Africa. AIDS 11, 493–498.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1571–1574.

Roques, P., Robertson, D., Souquiere, S., Apetrei, C., Nerrienet, E., Barre-Sinoussi, F., Muller-Trutwin, M.C., Simon, F., 2004. Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. AIDS 18, 1371–1381.

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

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

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noe, R., Peeters, M., Brookfield, J.F., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2005. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J. Virol. 79, 12515–12527.

Sharp, P.M., Shaw, G.M., Hahn, B.H., 2005. Simian immunodeficiency virus infection of chimpanzees. J. Virol. 79, 3891–3902.

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

Simon, F., Souquiere, S., Damond, F., Kfutwah, A., Makuwa, M., Leroy, E., Rouquet, P., Berthier, J.L., Rigoulet, J., Lecu, A., Telfer, P.T., Pandrea, I., Plantier, J.C., Barre-Sinoussi, F., Roques, P., Muller-Trutwin, M.C., Apetrei, C., 2001. Synthetic peptide strategy for the detection of and discrimination among highly divergent primate lentiviruses. AIDS Res. Hum. Retroviruses 17, 937–952.

Takehisa, J., Kraus, M.H., Decker, J.M., Li, Y., Keele, B.F., Bibollet-Ruche, F., Zammit, K.P., Weng, Z., Santiago, M.L., Kamenya, S., Wilson, M.L., Pusey, A.E., Bailes, E., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2007. Generation of infectious molecular clones of simian immunodeficiency virus from fecal consensus sequences of wild chimpanzees. J. Virol. 81 (14), 7463–7475.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. 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.

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

Van de Woude, S., Apetrei, C., 2006. Going wild: lessons from naturally occurring T-lymphotropic lentiviruses. Clin. Microbiol. Rev. 19, 728–762.

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

Vergne, L., Bourgeois, A., Mpoudi-Ngole, E., Mougnutou, R., Mbuagbaw, J., Liegeois, F., Laurent, C., Butel, C., Zekeng, L., Delaporte, E., Peeters, M., 2003. Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage. Virology 310, 254–266.

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

Yamaguchi, J., Bodelle, P., Vallari, A.S., Coffey, R., McArthur, C.P., Schochetman, G., Devare, S.G., Brennan, C.A., 2004. HIV infections in northwestern Cameroon: identification of HIV type 1 group O and dual HIV type 1 group M and group O infections. AIDS Res. Hum. Retroviruses 20, 944–957.

Yamaguchi, J., Coffey, R., Vallari, A.S., Ngansop, C., Mbanya, D., Ndembi, N., Kaptue, L., Gurtler, L.G., Bodelle, P., Schochetman, G., Devare, S.G., Brennan, C.A., 2006. Identification of HIV type 1 group N infections in a husband and wife in Cameroon: viral genome sequences provide evidence for horizontal transmission. AIDS Res. Hum. Retroviruses 22, 83–92.

Yang, Z., Rannala, B., 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. Mol. Biol. Evol. 14, 717–724.

Yang, C., Dash, B.C., Simon, F., van der groen, G., Pieniazek, D., Gao, F., Hahn, B.H., Lal, R.B., 2000. Detection of diverse variants of human immunodeficiency virus-1 groups M, N, and O and simian immunodeficiency viruses from chimpanzees by using generic *pol* and *env* primer pairs. J. Infect. Dis. 181, 1791–1795.