Reference Sequences Representing the Principal Genetic Diversity of HIV-1 in the Pandemic

Jean K. Carr,[1] **Brian T. Foley,**[2] **Thomas Leitner,**[3] **Mika Salminen,**[4] **Bette Korber,**[2] **and Francine McCutchan**[1]

[1] *Henry M. Jackson Foundation, 1 Taft Court, Rockville, MD 20850, USA*
[2] *Theoretical Biology and Biophysics (T10), MS K710, Los Almos National Laboratory, Los Alamos, NM 87545, USA*
[3] *Department of Virology, Swedish Institute for Infectious Disease Control, Karolinska Institute, S-171 82 Solna, Sweden*
[4] *National Public Health Institute, HIV Laboratory, Department of Infectious Diseases, Mannerheimintie 166, FIN-00300 Helsinki, Finland*

The phylogenetic complexity of the primate lentiviruses is legendary and our understanding of its scope was dramatically increased recently by the discovery of a new branch of the HIV-1 cluster, the "N" viruses [Simon *et al*., 1998]. There are now two major branches, "N" and "O", in the phylogenetic tree of HIV-1 sequences in addition to the cluster of sequences that form the "M", or Main group (Figure 1). The HIV-1 M, N, O, and the chimpanzee CPZ sequences cluster some what differently depending on the region compared (Figure 1). This is discussed in detail in conjunction with a new CPZ sequence, CPZ-US, a full length sequence described in [Gao *et al*., 1999]. (The CPZ-US sequence became available too late for inclusion in this study, but is available in the reference set at our website, http://hiv-web.lanl.gov/ALIGN_99/subtype_alignments.html.) At this point in time, however, only the viruses in the M group have significant public health importance. Genetic diversity within the M group takes the form of phylogenetic clusters which have been named subtypes. There are now at least 8 different subtypes of HIV-1 which circulate to varying extents in populations around the globe. A variety of factors make the genetic structure of HIV-1 particularly fluid both in time and space. This article will provide a description of our current understanding of the major circulating forms in the HIV-1 epidemic, and a subtyping reference set which can be used as a basis for the classification of new sequences.

The number of HIV-1 viruses which have been sequenced in their entirety has increased dramatically in the past few years [Korber *et al*., 1997], as have the number of tools designed to detect the presence of mosaic genomes [Salminen *et al*., 1995; Siepel *et al*., 1995]. It is important to distinguish newly discovered subtypes from recombinants, and to identify recombinant forms of epidemic importance. Now that full-length genomic sequencing is no longer a major obstacle, we propose that a virus isolate should fulfill the following criteria to be considered a subtype: (1) at least two isolates should be sequenced in their entirety, (2) they should resemble each other but no other existing subtype throughout the genome and, (3) they should have been found in at least two epidemiologically unlinked individuals. By these criteria, there are currently 8 subtypes of HIV-1. We are also aware that there are many mosaic genomes of HIV-1, some of which are unique, or restricted to one isolated transmission cluster, and others which are major circulating forms. Recombinant viruses are not as uncommon as previously thought and are especially prevalent in populations where multiple subtypes co-circulate. While possibly interesting for other reasons, the unique recombinant viruses do not play a major role in the global epidemic. In contrast, mosaic viruses which have spread from one location to another and have been associated with new outbreaks of the virus, such as the AE recombinant virus in Southeast Asia, have established a distinct and recognizable genetic lineage. It is proposed that those recombinant viruses be designated "Circulating Recombinant Forms", or CRFs, in distinction to the recombinants which are not known to be in circulation. We propose that a virus isolate should fulfill the following criteria to be considered a CRF: (1) at least two isolates should be sequenced in their entirety, (2) they should resemble each other but no other existing CRF in their subtype structure and (3) they should have been found in at least two epidemiologically unlinked individuals. These forms can be distinguished by associating the CRF with the name of the first full-length viral sequence of

that form. By these criteria, there are currently 4 CRFs of HIV-1, the AE virus from Southeast Asia, called "AE(CM240)" [Carr *et al*., 1996; Gao *et al*., 1996a; Gao *et al*., 1996b], the AG recombinant from west and central Africa, called "AG(IbNG)" [Carr *et al*., 1998], the AGI recombinant from Cyprus and Greece, called "AGI(CY032)" [Gao et al., 1998a; Kostrikis *et al*., 1995; Nasioulas *et al*., 1999], and the AB recombinant from Russia, called "AB(Kal153)" This sequence was provided prior to publication by Mika Salminen and is representative of the CRF found in the Kaliningrad IVDU epidemic described in [Liitsola *et al*., 1998]. The 8 subtypes and 4 major circulating recombinant forms create 12 major branches in the phylogenetic tree representing the lineage of the M group of HIV-1 (Figure 1).

The 12 major genetic forms of the HIV-1 M group are listed, with selected full-length sequences to use as references, in Table 1. The subtypes are A, B, C, D, F, G, H, J, and the four CRFs are: AE(CM240), AG(IbNG), AGI(CY032) and AB(KAL153). New to this edition of the database are the first full-length sequences from subtypes F, G, H and J, as well as more new isolates from subtypes A, C and D. In addition, there are new full-length sequences from the CRFs AGI(CY032), AB(Kal153) and AG(IbNG).

Subtype A, the most prevalent subtype in Africa, has recombined with many other subtypes in a myriad of permutations and combinations. So far, however, only four of those recombinations, to our knowledge, have yielded viruses which have spread to a significant extent. The first is a recombination with subtype E, forming the AE virus prevalent in Southeast Asia [Carr *et al*., 1996; Gao et al., 1996a; Gao *et al*., 1996b]. The parental E virus has never been found. The virus contains a subtype E env and LTR, but most if not all of the remainder of the virus derives from subtype A. The second and third A recombinants which have spread extensively are ones which have recombined with subtype G. The first of these viruses is called "AG(IbNG)". The first isolate of this CRF which was fully sequenced was from Ibadan, Nigeria and was named "IbNG" [Howard *et al*., 1994; Howard *et al*., 1996; Gao et al., 1996b]. Other viruses with the same structure have been fully sequenced from Djibouti and Ivory Coast [Carr *et al*., 1998; Carr *et al*., 1999] and there are many partial sequences from west or west central Africa, all of which cluster with IbNG [Ellenberger *et al*., 1999; Takehisa *et al*., 1998; McCutchan *et al*., 1999]. The AG(IbNG) virus is mosaic in pol and LTR, but since both gag and env derive largely from subtype A these viruses were initially classified as subtype A [Howard *et al*., 1996]. In fact, in both gag and env they form a significant subcluster within the A subtype and can be recognized even using partial sequencing of familiar regions. The third major CRF is AGI(CY032), a recombinant between subtypes A and G and possibly another previously unknown subtype, I. Like the parental E virus, the parental, "pure" I virus is not known. Two of the three viruses of this type have been found in epidemiologically unlinked individuals in Cyprus and Greece [Gao *et al*., 1998; Nasioulas *et al*., 1999]. The last A recombinant to be identified is the AB(Kal153) virus from the city of Kaliningrad in Russia. Some of this recombinant is from subtype A but most of the env region is subtype B. This recombinant has been responsible for an explosive epidemic among drug users in the city of Kaliningrad [Liitsola *et al*., 1998].

The genetic structure of the first full length AG(IbNG) and subtype G viruses has been recently published [Carr *et al*., 1998]. In protease, in the accessory gene region, and at the very end of env, an unusual phylogenetic relationship exists between subtypes A, G, AE(CM240) and AG(IbNG). In a genetic sense, they are neither as close, nor as distant, as in other regions of the genome, where it is simple to identify the region as belonging to a given subtype. In these regions they show an intermediate relationship. While this phenomenon is observed with the A, G, E and IbNG cluster, it is not observed with the other subtypes in the same regions. It is therefore not due to a general weakness in the information content of the region or to the analytic approach. Some have suggested that the G viruses are actually recombinant with subtype A in these regions (Gao *et al*., 1998), and while this is a possibility, others are unable to convincingly demonstrate a recombinant nature for the G viruses [Carr *et al*., 1998]. At the moment the issue is not completely resolved.

A variety of intersubtype recombinants combining segments of A and C, A and D, B and F, and others have been described or are known in yet-to-be-published studies. Each of these unique forms could potentially spread epidemically, and as new recombinants are studied it is increasingly important to compare them in detail to the full spectrum of known recombinant forms. The initial events leading to the emergence of recombinants may be better understood in future years.

Table 1A Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes and Circulating Recombinant Forms

Subtype	Sequence ID's				*Region*
A	U455	92UG037.1	Q2317	SE7253	complete genome
B	HXB2	JRFL	RF	WEAU.160	complete genome
C	ETH2220	92BR025	IN21068	96BW05.02	complete genome
D	NDK	ELI	94UG114[1]	84ZR085	complete genome
F	93BR020.1 BZ162 BZ163	FIN6393 VI69 BZ126	VI850		complete genome *gag* *env*
G[2]	SE6165	HH8793	DRCBL		complete genome
H	90CF056.1 VI557[3]	VI991	VI997		complete genome *gag*, *env*
J	SE9280.9	SE9173.3			complete genome
Circulating Recombinant Forms (CRFs)					
AB(KAL153)	KAL153[4]				complete genome
AE(CM240)	CM240[5] TN235	90CR402.1	93TH253.3		complete genome *env*
AG(IbNG)	IBNG[6]	DJ263	DJ264		complete genome
AGI(CY032)	CY032.3[7]	PVMY	PVCH		complete genome
Group					
N	YBF30				complete genome
O	MVP5180	ANT70			complete genome
Chimpanzee SIV Isolates					
CPZ	CPZANT[8]	CPZGAB			complete genome

[1] The sequence 94UG114.1 is the most distant complete genome D subtype sequence (see trees), tending to branch off closest to the B/D root in most analyses. In some subgenomic regions, it may even move outside the B/D cluster.

[2] The G reference sequences may show resemblance to subtype A in regions of *pol*, *vif*, and *env* [2, 8]. The sequence NG083.2, characterized as an AG in [8] shares a similar genetic structure with the three sequences listed as G here.

[3] Full length gene sequences of *gag*, *pol*, or *env* are not yet available, see Table 1B.

[4] A circulating recombinant form with A and B subtype subregions in an epidemic strain in an outbreak in Kaliningrad. Only one full-length genome is available at this time [14].

[5] Circulating Recombinant Form AE was called subtype E in previous compendia and the previous literature, but it is E in most of *env*, A in *gag* and *pol*, mixture of A & E in regulatory genes [3, 8, 14].

[6] Circulating Recombinant Form AG(IBNG) recombination breakpoints are described in [2].

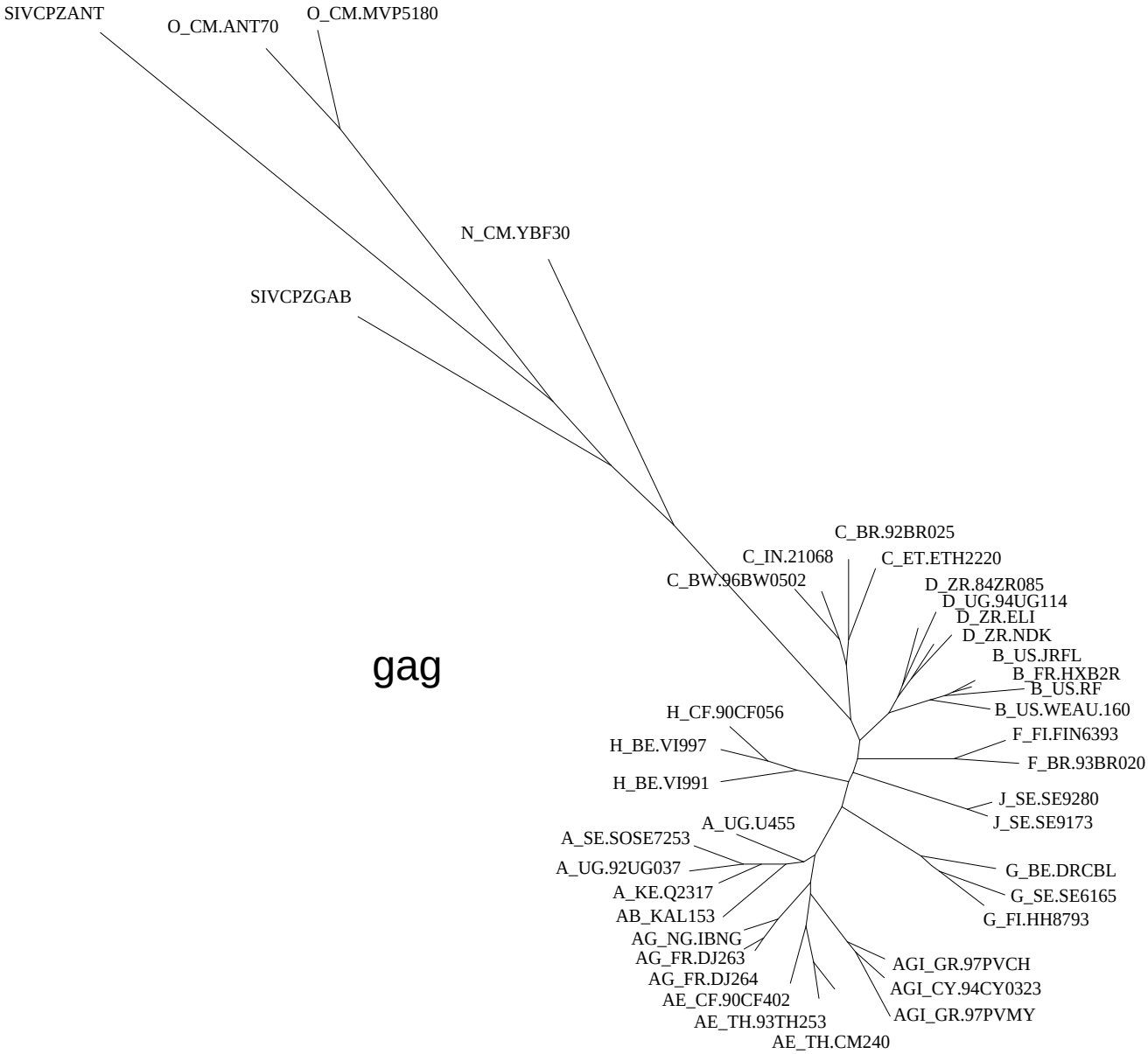
[7] Circulating Recombinant Form AGI(CY032) recombination breakpoints are described in [7].

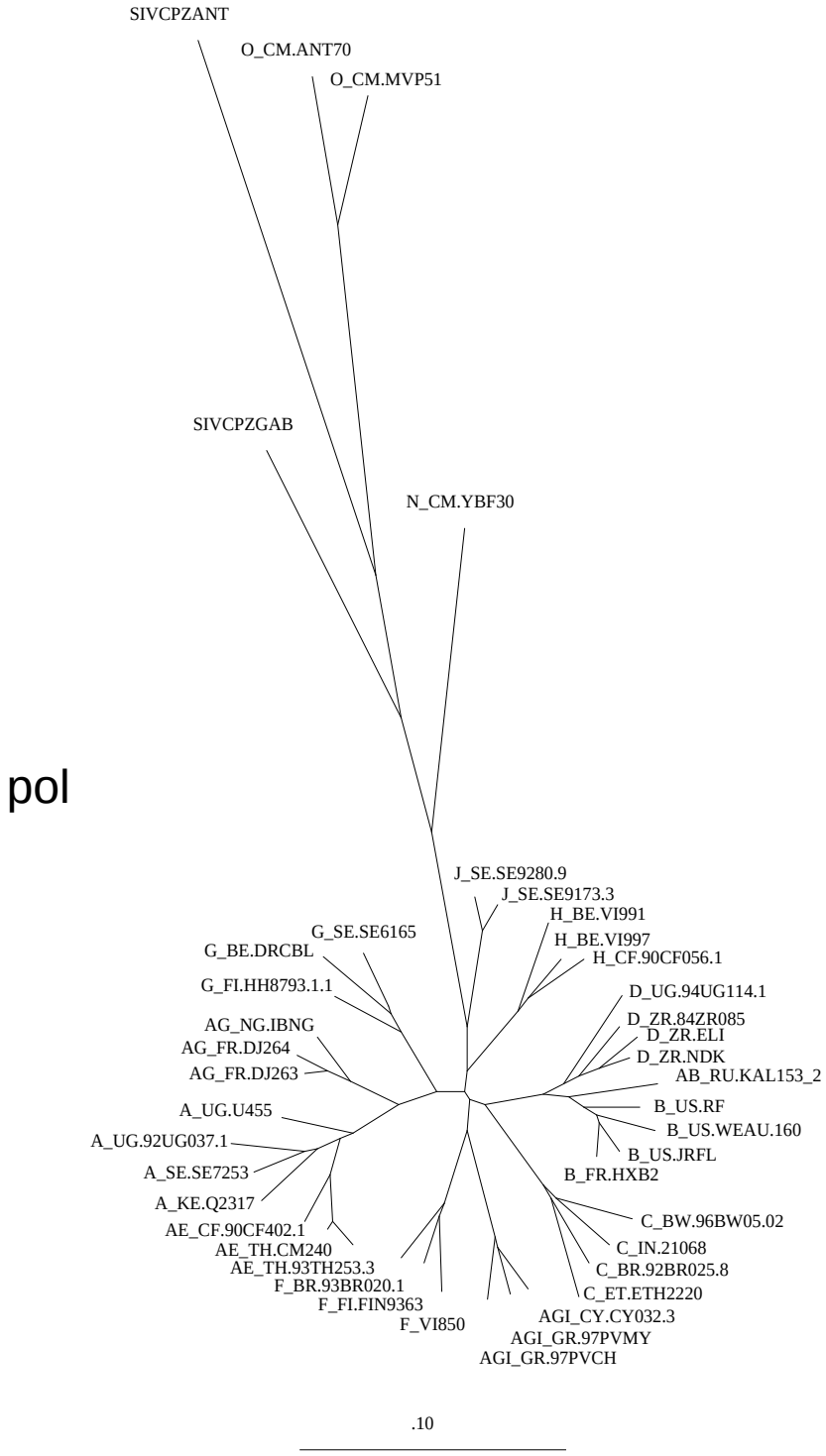
[8] A 275 bp insert in CPZANT, near the 3′ end of the *pol* gene was deleted from the subtype reference set alignment, because this insert is not related to any HIV or SIV sequence, and likely represents a cloning or sequencing artifact. See pages I–54 to I–55 of the 1997 HIV Sequence Compendium.

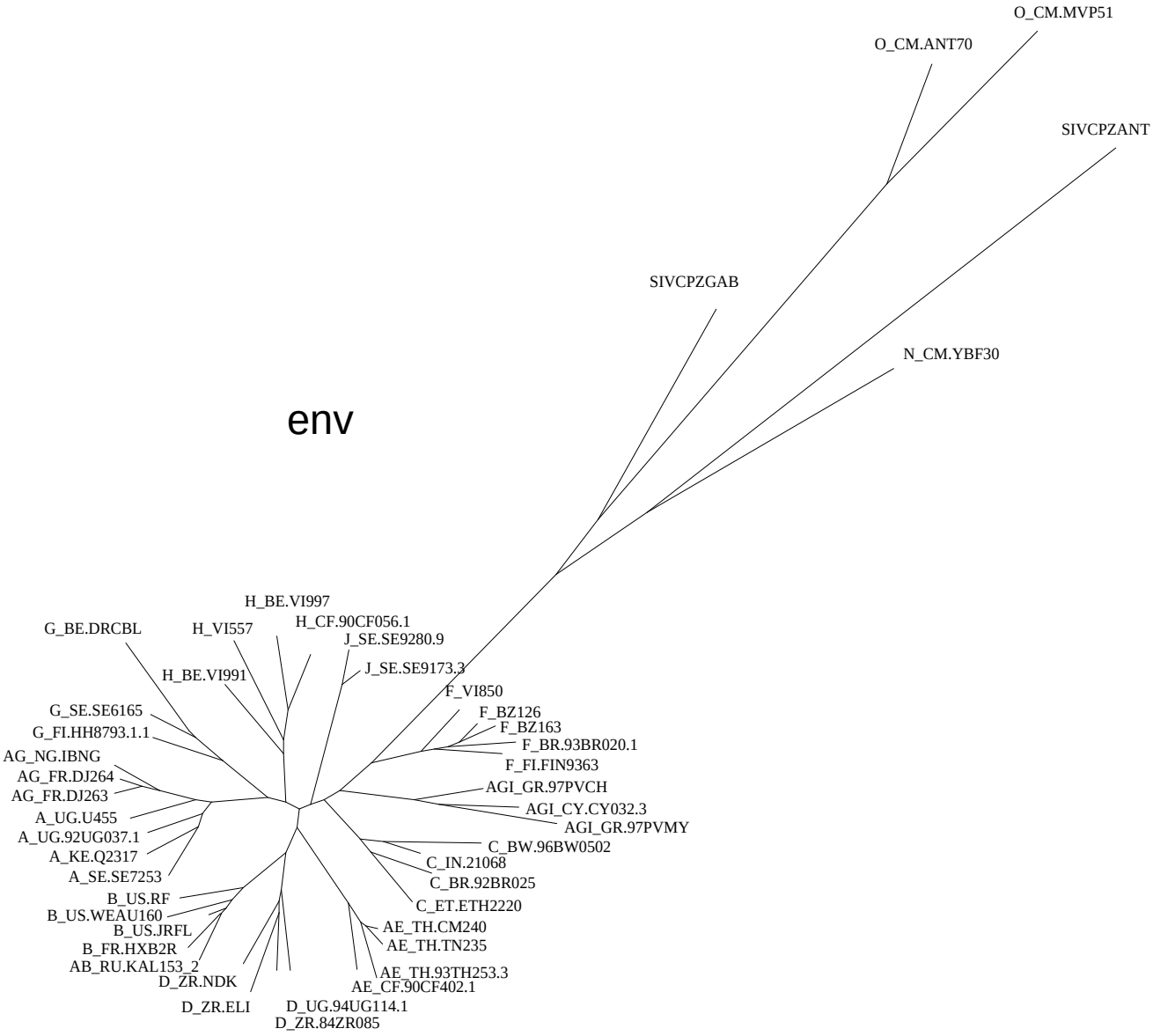
Table 1B Description of sequences in alignments

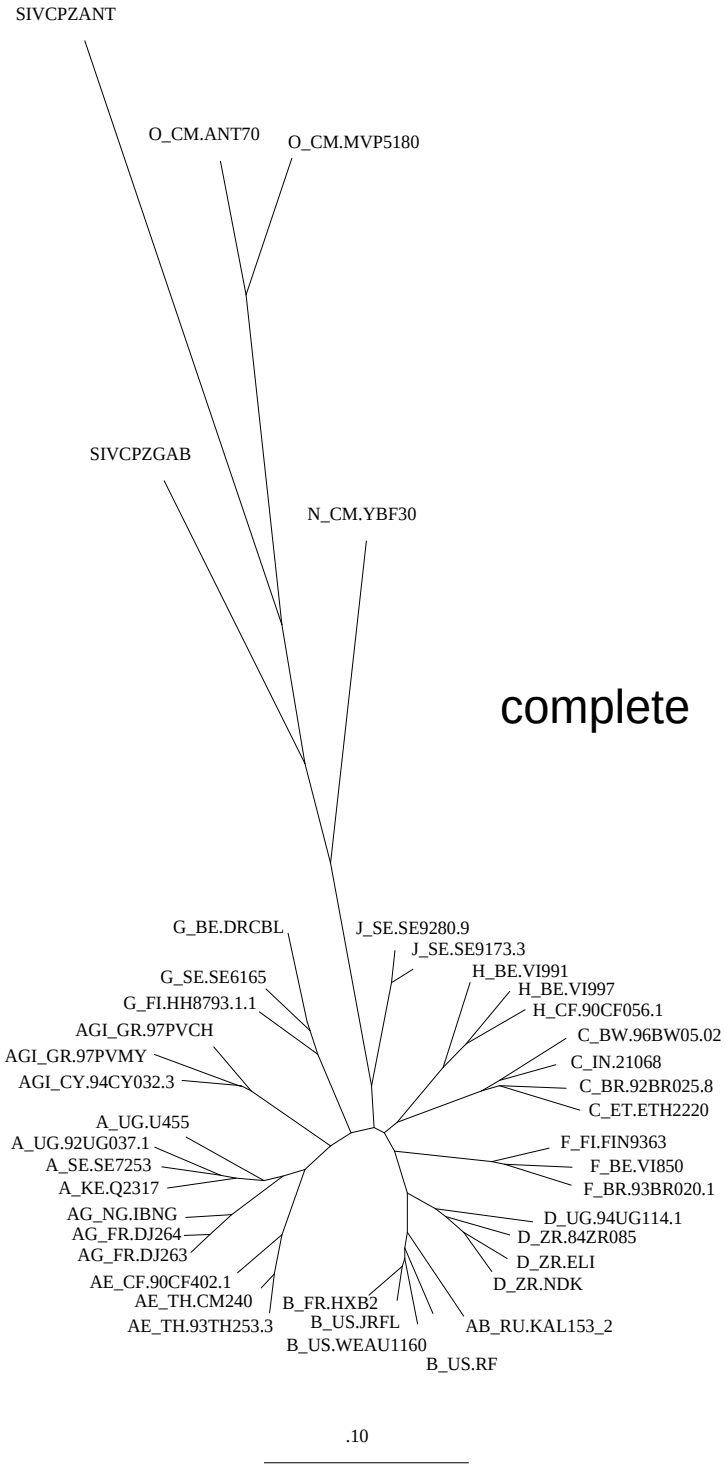
Subtype	Sequence	Acc. No.	Source	Region	Sampling year	Sampling country (origin)
A	U455	M62320	Oram, J.D. et al, *ARHR* **6**:1073–1078 (1990)	complete	NA	Uganda
A	92UG037.1	U51190	Gao, F. et al, *J Virol* **70**:7013–7029 (1996)	complete	1992	Uganda
A	Q2317	AF004885	Poss, M. et al, *J Virol* **72**:8240–8251 (1998)	complete	1994	Kenya
A	SE7253	AF069670	Carr, J.K. et al, *Unpublished*. (1998)	complete	1994	Sweden (Somalia)
B	HXB2	K03455, M38432	Wong-Staal, F. et al, *Nature* **313**:277–284 (1985)	complete	NA	France
B	JRFL	U63632	O'Brien, W.A. et al, *Nature* **348**:69 (1990)	complete	NA	U.S.A.
B	RF	M17451, M12508	Starcich, B.R. et al, *Cell* **45**:637–648 (1986)	complete	1983	U.S.A. (Haiti)
B	WEAU160	U21135	Ghosh, S.K. et al, *Unpublished* (1995)	complete	NA	U.S.A.
C	ETH2220	U46016	Salminen, M.O. et al, *ARHR* **12**:1329–1339 (1996)	complete	1986	Ethopia
C	92BR025.8	U52953	Gao, F. et al, *J Virol* **70**:1651–1657 (1996)	complete	1992	Brazil
C	IN21068	AF067155	Lole, K.S. et al, *Unpublished* (1998)	complete	NA	India
C	96BW05.02	AF110967	Novitsky, V. et al, *J Virol* submitted (1998)	complete	1996	Botswana
D	NDK	M27323	Spire, B. et al, *Gene* **81**:275–284 (1989)	complete	NA	Zaire/DRC*
D	ELI	K03454, X04414	Alizon, M. et al, *Cell* **46**:63–74 (1986)	complete	NA	Zaire/DRC*
D	94UG114.1	U88824	Gao, F. et al, *J Virol* **72**:5680–5698 (1998)	complete	1994	Uganda
D	84ZR085	U88822	Gao, F. et al, *J Virol* **72**:5680–5698 (1998)	complete	1984	Zaire/DRC*
F	93BR020.1	AF005494	Gao, F. et al, *J Virol* **72**:5680–5698 (1998)	complete	1993	Brazil
F	VI850	AF077336	Carr, J. et al., *Unpublished* (1998)	complete	1993	Belgium (Zaire/DRC*)
F	FIN9363	AF075703	Laukkanen, T. et al, *Unpublished* (1998)	complete	NA	Finland
F	BZ162	L11751	Louwagie, J. et al, *AIDS* **7**:769–780 (1993)	*gag*	NA	Brazil
F	VI69	L11796	Louwagie, J. et al, *AIDS* **7**:769–780 (1993)	*gag*	NA	Belgium (Rwanda)
F	BZ163	L22085	Louwagie, J. et al, *ARHR* **10**:561–567 (1994)	*env-nef*	NA	Brazil
F	BZ126	L22082	Louwagie, J. et al, *ARHR* **10**:561–567 (1994)	*env-nef*	NA	Brazil
G**	SE6165	AF061642	Carr, J.K. et al, *Virology* **247**:22–31 (1998)	complete	1993	Sweden (Zaire/DRC*)
G**	HH8793.1.1	AF061640	Carr, J.K. et al, *Virology* **247**:22–31 (1998)	complete	NA	Finland (Kenya)
G**	DRCBL	AF084936	Oelrichs, R. et al, *ARHR* in press (1999)	complete	NA	Belgium
H	90CF056.1	AF005496	Gao, F. et al, *J Virol* **72**:5680–5698 (1998)	complete	1990	Cent. Afr. Rep.
H	VI991	none yet	Laukkanen, T. et al, *Unpublished* (1998)	complete	NA	Belgium
H	VI997	none yet	Laukkanen, T. et al, *Unpublished* (1998)	complete	NA	Belgium
H	VI557	U09666	Janssens, W. et al, *ARHR* **10**:877–879 (1994)	V3-V5	NA	Zaire/DRC*
H	VI557	L11793	Louwagie, J. et al, *AIDS* **7**:769–780 (1993)	*gag*	NA	Zaire/DRC*
J	SE9280.9	AF082394	Laukkanen, T. et al, *ARHR* in press (1999)	complete	NA	Sweden
J	SE9173.3	AF082395	Laukkanen, T. et al, *ARHR* in press (1999)	complete	NA	Sweden (Zaire/DRC*)
AB	KAL153	none yet	Salminen, M.O., *Unpublished* (1998)	complete	NA	Russia
AE	CM240	U54771	Carr, J.K. et al, *J Virol* **70**:5935–5943 (1996)	complete	1990	Thailand
AE(like CM240)	90CF402.1	U51188	Gao, F. et al, *J Virol* **70**: 7013–7029 (1996)	complete	1990	Cent. Afr. Rep.
AE(like CM240)	93TH253.3	U51189	Gao, F. et al, *J Virol* **70**: 7013–7029 (1996)	complete	1993	Thailand
AE(like CM240)	TN235	L03698	McCutchan, F.E. et al, *ARHR* **8**:1887–1895 (1992)	*env*	NA	Thailand
AG	IBNG	L39106	Howard, T.M. et al, *ARHR* **10**:1755–1757 (1994)	complete	NA	Nigeria
AG(like IBNG)	DJ264	AF063224	Carr, J.K. et al, *Virology* **247**:22–31 (1998)	complete	NA	France (Djibouti)
AG(like IBNG)	DJ263	AF063223	Carr, J.K. et al, *Virology* **247**:22–31 (1998)	complete	NA	France (Djibouti)
AGI	CY032.3	AF049337	Gao, F. et al, *J Virol* in press (1998)	complete	1994	Cyprus (Greece)
AGI(like CY032)	PVMY	AF119819	Nasioulas, G. et al, *ARHR* in press (1999)	complete	NA	Greece
AGI(like CY032)	PVCH	AF119820	Nasioulas, G. et al, *ARHR* in press (1999)	complete	NA	Greece
N	YBF30	AJ006022	Simon, F. et al, *Nature Medicine* **4**:1032–1037 (1998)	complete	1995	Cameroon
O	MVP5180	L20571	Gurtler, L.G. et al, *J Virol* **68**:1581–1585 (1994)	complete	1991	Cameroon
O	ANT70	L20587	Haesevelde, M. et al, *J Virol* **68**:1586–1596 (1994)	complete	NA	Cameroon
CPZ	CPZGAB	X52154	Huet, T. et al, *Nature* **345**:356–359 (1990)	complete	NA	Gabon
CPZ	CPZANT	U42720	Haesevelde, M. et al, *Virol* **221**:346–350 (1996)	complete	NA	Zaire/DRC*

* The former Zaire is now called the Democratic Republic of Congo
** These G viruses are difficult to classify, as they contain some possibly "A" like regions.


SIVCPZANT
O_CM.ANT70
O_CM.MVP5180
N_CM.YBF30
SIVCPZGAB
gag
C_BR.92BR025
C_IN.21068
C_ET.ETH2220
C_BW.96BW0502
D_ZR.84ZR085
D_UG.94UG114
D_ZR.ELI
D_ZR.NDK
B_US.JRFL
B_FR.HXB2R
B_US.RF
B_US.WEAU.160
H_CF.90CF056
H_BE.VI997
H_BE.VI991
F_FI.FIN6393
F_BR.93BR020
J_SE.SE9280
J_SE.SE9173
A_UG.U455
A_SE.SOSE7253
A_UG.92UG037
A_KE.Q2317
AB_KAL153
AG_NG.IBNG
AG_FR.DJ263
AG_FR.DJ264
AE_CF.90CF402
AE_TH.93TH253
AE_TH.CM240
G_BE.DRCBL
G_SE.SE6165
G_FI.HH8793
AGI_GR.97PVCH
AGI_CY.94CY0323
AGI_GR.97PVMY
.10


SIVCPZANT
O_CM.ANT70
O_CM.MVP51
SIVCPZGAB
N_CM.YBF30
pol
J_SE.SE9280.9
J_SE.SE9173.3
H_BE.VI991
H_BE.VI997
H_CF.90CF056.1
G_SE.SE6165
G_BE.DRCBL
G_FI.HH8793.1.1
AG_NG.IBNG
AG_FR.DJ264
AG_FR.DJ263
A_UG.U455
A_UG.92UG037.1
A_SE.SE7253
A_KE.Q2317
AE_CF.90CF402.1
AE_TH.CM240
AE_TH.93TH253.3
F_BR.93BR020.1
F_FI.FIN9363
F_VI850
D_UG.94UG114.1
D_ZR.84ZR085
D_ZR.ELI
D_ZR.NDK
AB_RU.KAL153_2
B_US.RF
B_US.WEAU.160
B_US.JRFL
B_FR.HXB2
C_BW.96BW05.02
C_IN.21068
C_BR.92BR025.8
C_ET.ETH2220
AGI_CY.CY032.3
AGI_GR.97PVMY
AGI_GR.97PVCH
.10


O_CM.MVP51
O_CM.ANT70
SIVCPZANT
SIVCPZGAB
N_CM.YBF30
env
H_BE.VI997
H_CF.90CF056.1
G_BE.DRCBL
H_VI557
J_SE.SE9280.9
J_SE.SE9173.3
H_BE.VI991
F_VI850
G_SE.SE6165
F_BZ126
F_BZ163
G_FI.HH8793.1.1
F_BR.93BR020.1
AG_NG.IBNG
F_FI.FIN9363
AG_FR.DJ264
AGI_GR.97PVCH
AG_FR.DJ263
AGI_CY.CY032.3
A_UG.U455
AGI_GR.97PVMY
A_UG.92UG037.1
C_BW.96BW0502
A_KE.Q2317
C_IN.21068
A_SE.SE7253
C_BR.92BR025
B_US.RF
C_ET.ETH2220
B_US.WEAU160
AE_TH.CM240
B_US.JRFL
AE_TH.TN235
B_FR.HXB2R
AE_TH.93TH253.3
AB_RU.KAL153_2
AE_CF.90CF402.1
D_ZR.NDK
D_ZR.ELI
D_UG.94UG114.1
D_ZR.84ZR085
.10


SIVCPZANT
O_CM.ANT70
O_CM.MVP5180
SIVCPZGAB
N_CM.YBF30
complete
G_BE.DRCBL
J_SE.SE9280.9
J_SE.SE9173.3
H_BE.VI991
G_SE.SE6165
H_BE.VI997
G_FI.HH8793.1.1
H_CF.90CF056.1
AGI_GR.97PVCH
C_BW.96BW05.02
AGI_GR.97PVMY
C_IN.21068
AGI_CY.94CY032.3
C_BR.92BR025.8
C_ET.ETH2220
A_UG.U455
A_UG.92UG037.1
F_FI.FIN9363
A_SE.SE7253
F_BE.VI850
A_KE.Q2317
F_BR.93BR020.1
AG_NG.IBNG
D_UG.94UG114.1
AG_FR.DJ264
D_ZR.84ZR085
AG_FR.DJ263
D_ZR.ELI
AE_CF.90CF402.1
D_ZR.NDK
AE_TH.CM240
B_FR.HXB2
AE_TH.93TH253.3
B_US.JRFL
AB_RU.KAL153_2
B_US.WEAU1160
B_US.RF
.10

REFERENCES

[1] Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol*. **70**:5935–5943.

[2] Carr, J. K., M. O. Salminen, J. Albert, E. Sanders-Buell, D. Gotte, D. L. Birx, and F. E. McCutchan. 1998 Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* **247**(1):22–31

[3] Carr, J.K., T. Laukkanen, M. O. Salminen, J. Albert, A. Alaeus, B. Kim, E. Sanders-Buell, D. L. Birx and F. E. McCutchan. 1999. Genetic Characterization of HIV-1 Subtype A Full Length Genomes from Africa. Submitted.

[4] Ellenberger, D.L., D. Pieniazek, J. Nkengasong, C.-C. Luo, S. Devare, C. Maurice, M. Janini, A. Ramos, C. Fridlund, D. J. Hu, I.-M. Coulibaly, E. Ekpini, S. Z. Wiktor, A. E. Greenberg, G. Schochetman and M. A. Rayfield. 1999. Genetic Analysis of Human Immunodeficiency Virus in Abidjan, Ivory Coast reveals predominance of HIV type 1 subtype A and introduction of subtype G. *AIDS Res Hum Retroviruses* **15**:3-9.

[5] Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, B. H. Hahn, and the WHO and NIAID networks for HIV isolation and characterization. 1996a. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol*. **70**(3): 1651–1667.

[6] Gao F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker J, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn and P. M. Sharp. 1996b. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol*. **70**(10):7013–1029.

[7] Gao F., D. L. Robertson, C. D. Carruthers, Y. Li, E. Bailes, L. G. Kostrikis, M. O. Salminen, F. Bibollet-Ruche, M. Peeters, D. D. Ho, G. M. Shaw, P. M. Sharp and B. H. Hahn. 1998a. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol*. **72**(12): 10234–10241

[8] Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barre- Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998b. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol*. **72**(7): 5680–5698

[9] Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp and B. H. Hahn. 1999. Origin of HIV-1 in *Pan troglodytes troglodytes*. *Nature*. **397**(6718): 436-441

[10] Howard, T. M., D. O. Olaylele and S. Rasheed. 1994. Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria. *AIDS Res. Hum. Retrovirus*. **10**(12): 1755–1757.

[11] Howard T. M. and S. Rasheed 1996 Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *AIDS Res. Hum. Retrovirus*. **12**(15):1413–1425.

[12] Korber, B., B. Hahn, B. Foley, J. W. Mellors, T. Leitner, G. Myers, F. McCutchan and C.L. Kuiken (ed.) *Human Retroviruses and AIDS 1997: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos National Laboratory, Los Alamos, NM.

[13] Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol*. **69**:6122–6130.

[14] Liitsola K., I. Tashkinova, T. Laukkanen, G. Korovina, T. Smolskaja, O. Momot, N. Mashkilleyson, S. Chaplinskas, H. Brummer-Korvenkontio, J. Vanhatalo, P. Leinikki and M. O. Salminen. 1998. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* **12**(14):1907–1919.

[15] McCutchan, F. E., J. K. Carr, M. Bajani, E. Sanders-Buell, T. Harry, T. C. Stoeckli, K. E. Robbins, W. Gashau, A. Nasidi, W. Janssens and M. L. Kalish. 1999. Subtype G and multiple forms of A/G interspecific recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* in press.

[16] Nasioulas, G., D. Paraskevis E. Magiorkinis, M. Theodoridou and A. Hatzakis. 1999. *AIDS Res Hum Retroviruses* in press

[17] Salminen, M.O., Carr, J.K., Burke, D.S. and F.E. McCutchan. 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995 **11**(11):1423–1425

[18] Siepel, A.C., Halpern, A.L., Macken, C. and B.T. Korber. 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. 1995. *AIDS Res Hum Retroviruses* **11**(11): 1413–1416

[19] Simon, F., P. Mauclere, P. Roques, I. Loussert-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi and F. Brun-Vezinet. 1998 Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**(9):1032–1037

[20] Takehisa, J., L. Zekeng, E. Ido, I. Mboudjeka, H. Moriyama, T. Miura, M. Yamashita, L. G. Gurtler, M. Hayami and L. Kaptue. 1998. Various types of HIV mixed infections in Cameroon. *Virology* **245**(1):1-10.