Table 2 Notes on selected full-length nucleotide sequences that appear in the alignments.

Sequence	Accession	Origin	Reference
B_FR.HXB2R	K03455	France	Wong-Staal F., *Nature* **313**:277–284 (1985)

HXB2R is the reference strain for this alignment. This sequence was from provirus cloned in lambda phage, and is derived from the IIIB isolate related to the French isolate LAI (Wain-Hobson, S., *Cell* **40**:9–17 (1985)). There are several other LAI-IIIB related clones that have been fully sequenced (LAI/BRU, BH10, PV22, PM213, MCK1, LW12–3, HXB2, and TH475), but in our printed compendium we include only the common reference strain HXB2. The alignments on our web site (http://hiv-web.lanl.gov/ALIGN_99/ALIGN-INDEX-99.html) and ftp site include the full set of full length LAI-related HIV sequences. These were the first HIV isolates, are T-cell tropic, and are the most extensively studied variants in terms of mutational analysis of functional domains, phenotype, vaccine design, and antigenicity. The HXB2R clone sequence is now our standard reference for noting positions in the HIV-1 genome or proteins, as discussed in the article "Numbering Positions in HIV Relative to HXB2CG," in Part III of this compendium. We have also created a Web site (http://hiv-web.lanl.gov/NUM-HXB2/HXB2.MAIN.html) that automatically shows the position of your sequence relative to HXB2R. It was selected as the standard because it has been extensively studied, and many of its proteins crystallized. Both Tat and Nef have premature stop codons. As the use of LAI-related cultures, reagents and clones is ubiquitous in AIDS research, these sequences are occasionally found as contaminants in sequencing studies. (Although LAI/IIIB sequences are certainly not the only possible source of contamination, they are the most common).

A_KE.Q2317 AF004885 Kenya Poss, M. Unpublished (1997)

This subtype A sequence was derived from a woman from Mombasa, Kenya, who had been recently infected with HIV-1. The blood sample was drawn on June 13, 1994. An env gene fragment from a PCR amplification from an earlier blood sample (July 1993) was published in Poss M., *ARHR* **13**:493–499 (1997). The full length sequence was kindly released prior to publication by M. Poss and colleagues, U. Washington. Many env sequences from this same patient are available with accession numbers AF004893 and AF047979–AF048346.

A_SE.SOSE7253 AF069670 Sweden(Somalia) Carr, J.K. Unpublished (1998)

This subtype A sequence is from a 27 year old male living in Sweden, who is thought to have been infected in Somalia via heterosexual contact. The patient was CDC stage C3 when sampled in 1994. The virus is NSI and uses the CCR5 coreceptor. The patient's CD4 count was zero. Virus was cocultured with donor PBMC before PCR amplification and direct sequencing. Small sections of the 5′ and 3′ LTRs are not included in this sequence.

A_UG.92UG037.1 U51190 Uganda Gao, F. J Virol **72**(7), 5680–98 (1998)

Sample 92UG037 is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, *ARHR* **10**:1327–1344 (1994)) and comes from an asymptomatic 31-year-old female from Entebbe, Uganda; she had not taken any anti-retroviral therapy prior to sampling. The risk factor for infection was heterosexual contact. The isolate 92UG037 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. 92UG037 is subtype A. An LTR sequence is available under accession numbers U51287, and an additional env/nef sequence has accession number U09127. There is an inframe stop codon in pol at position 3144 in this clone. The isolate from which this sequence was derived is NSI and uses CCR5 or CCR8 (Bjorndal, A. *J Virol* **71**:7478 (1997) and Rucker, J., *J Virol* **71**:8999–9007 (1997). See also: Gao, F. *J Virol* **70**:7013–7029 (1996).

Table 2 (cont.)

Sequence	Accession	Origin	Reference
A_UG.U455	M62320	Uganda	Oram, J. *ARHR* **6**:1073–1078 (1990)

This sequence is from the 1985 Ugandan isolate U455. It was cloned in phage, and is defective in env, vpr and vpu. The env ORF in this sequence is interrupted by an in-frame stop codon beyond the COOH end of the V5 region. This sequence clusters with subtype A HIV-1.

B_US.JRFL U63632 USA O'Brien, W. A., *Nature* **348**:69–73 (1990)

This sequence is from an infectious lambda phage clone of the 1986 isolate JRFL, derived from the frontal lobe of the brain of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy in 1996. The infectious clone JRCSF was isolated from the CSF of the same patient. Both are NSI. Also see: Pang, J., *JAIDS* **4**:1082–92 (1991) and Klasse, P. J., *ARHR* **12**:347–350 (1996).

B_AU.MBC200 AF042100 Oelrichs, R.B. Unpublished (1998)

This sequence is from an Australian man with AIDS, whose blood was drawn March 18, 1986. The patient was a Caucasian homosexual male, diagnosed with AIDS in December 1985 at which time the T4:T8 ratio was 0.2. The virus was biological cloned by three rounds of limiting dilution in donor PBMC. The sequence was derived directly from the biologically cloned isolate. Hirt supernatant DNA was obtained from low-passage number donor PBMC culture and sequence derived from overlapping PCR products. All open reading frames are intact. Viral phenotype: Produces high levels of syncytia in PBMC and MT-2 cells. Grows well in Jurkat cells and primary macrophages (Kiernan,R. *ARHR* **6**(6), 743–752 1990)).

B_US.MN M17449 USA Gurgo C., *Virology* **164**:531–536 (1988)

MN is from one of the earliest available isolates, and is a commonly used reference and vaccine strain. The MN isolate was taken from a 6 year old male pediatric AIDS patient in 1984, in Newark, New Jersy. The MN sequence was cloned from the isolate in lambda phage. The coding sequences for pol, nef and vpu are prematurely truncated. A set of V3 sequences from this isolate are available L48364-L48379, Lukashov, V. and Goudsmit J., *AIDS* **9**:1307–1311 (1995)

B_US.NY5CG M38431 USA Willey, R., *PNAS* **83**:5038–5042 (1986)

This sequence is from the 1984 New York T-cell tropic isolate NY5. It was cloned in lambda phage and is not replication competent. It has a defective vpu gene due to the loss of the start codon. See also GenBank accession number K03346, for an env gene sequence from this isolate.

B_NL.ACH320A U34604 The Netherlands Guillon, C., *ARHR* **11**:1537–1538 (1995)

This is the sequence of the complete genome of an NSI, macrophage tropic lambda phage clone from an isolate taken from the PBMC of a patient who was making a shift from NSI to SI in bulk phenotype. A syncytium inducing clone from the same isolate has also been completely sequenced, ACH320.2A.1.2. The patient, isolates and phenotype of the molecular clones are described in Groenink, M., *J Virol* **65**:1968–1975 (1991).

B_US.SF2CG K02007 USA Sanchez-Pescador, R., *Science* **227**:484–492 (1985)

This sequence is from an infectious phage clone from the U.S. isolate ARV-2. ARV-2/SF2 was isolated from the PBMC of a patient with oral candidasis after co-culture with mitogen-stimulated PBMCs, (Levy, J., *Science* **225**:840–842, (1984)). It is a standard reference strain, and has been used for vaccine studies. SF2CG stands for SF2 complete genome.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
B_US.WEAU160	U21135	USA	Ghosh, S. K., Unpublished (1995)

The sequence is from a cytopathic HIV-1 virus clone from an isolate from an acutely infected patient, from Birmingham, Alabama. The lambda phage clone was obtained from a co-culture of this patient's PBMCs, first with normal donor PHA-stimulated lymphocytes for 14 days, and then with the H9 T-cell line for another 14 days. The blood specimen was obtained 15 days after the onset of clinical symptoms of acute (primary) infection, and 35 days after a single sexual encounter (receptive anal intercourse) with a partner whose virus was proven phylogenetically to be responsible for the transmission event. The clone that was sequenced has an SI phenotype, and a frameshift in nef. The single nucleotide deletion in nef found in the full length WEAU 1.60 clone is not present in 10/10 PCR amplified sequences from the patient's uncultured PBMCs (instead there is a "T".) The full-length WEAU 1.60 provirus was sequenced in its entirety by two different laboratories (George Shaw and Leroy Hood) with 100% concordance. George Shaw provided detailed information concerning this sequence. The patient WEAU is "Patient #1" in Clark, S. J. *N Engl J Med* **324**:954–960 (1991); "WEAU 0575" in Piatak, M., *Science* **259**:1749–1754 (1993) and is also discussed in Borrow, P., *Nat Med* **3**:205–11 (1997). WEAU was the reference strain for 1997 alignments as well as for the Los Alamos Molecular Immunology Database Compendium, the companion volume to the HIV sequence compendium.

B_US.RF M17451 USA Starcich, B. R., *Cell* **45**:637–648 (1986)

RF is among the first isolates, and is a commonly used reference and vaccine study strain. The sequence is from the full-length lambda phage clone HAT-3, from isolate RF, cultured in HUT-78 cells. RF is from a 28 year old symptomatic Haitian male, who moved to the U.S. in 1980 and was sampled in 1983, shortly before his death. RF has defective gag and vpu genes. Several env genes are available from this isolate, U30778-U30781. See also: Reitz M., *ARHR* **8**:1950 (1992)

B_CN.RL42 U71182 China Graf, M. *ARHR* 14(3):285–288 (1998)

RL42 was isolated from an asymptomatic IVDU, infected by needle sharing, in Dehong prefecture of Yunnan province of South China. This is near the Laos and Thailand golden drug triangle. The isolate was generated by Prof. Dr. Shao Yiming from the Chinese Academy of Preventive Medicine, Beijing, China. This sequence is of the Thai B′ subtype (a subset of subtype B), which is the most prevalent subtype of HIV-1 found in China.

C_BW.96BW0502 AF110967 Botswana Novitsky, V. Submitted (1998)

This subtype C sequence is from Botswana. It has four NF-κB binding sites where most subtype C have three, and most other subtypes have just two. It was kindly provided as a reference strain prior to publication by Dr. Vlad Novitsky, and is part of a study of multiple full length C subtype sequences from Botswana.

C_IN.11246 AF067159 India Lole, K.S. *J Virol* **73**(1):152–60 (1999)

This subtype C sequence is one of several complete genomes from India. It is derived from primary PBMC cocultures taken Oct. 25, 1994 from a 26 year old man from Pune in Maharashtra State, India, who seroconverted in 1994. His major risk factor for HIV infection was sexual contact with commercial sex workers. There is a termination codon at 251 bp downstream from the splice acceptor site found in all C subtype sequences from this study, C2220 and BR025. Clone 94IN11246 has an eight-base deletion at nucleotide 572 of nef, resulting in a frameshift and an open reading frame that is 136 bp longer than usual. This clone is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p94IN11246.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
C_IN.301904	AF067157	India	Lole, K.S. *J Virol* **73**(1):152–60 (1999)

This subtype C sequence is one of several complete genomes from India. It is derived from primary PBMC coculture taken March 27, 1993 from a 28 year old woman from Pune in Maharashtra State, India who seroconverted in 1992 following a blood transfusion; NSI phenotype. There is a termination codon at 251 bp downstream from the splice acceptor site found in all C subtype sequences from this study, C2220 and BR025. This clone is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p93IN301904.

C_IN.301905 AF067158 India Lole, K.S. *J Virol* **73**(1):152–60 (1999)

This subtype C sequence is one of several complete genomes from India. It is derived from primary PBMC coculture taken March 27, 1993 from a 23 year old woman from Pune in Maharashtra State, India who seroconverted in 1992 following a blood transfusion; NSI phenotype. There is a termination codon at 251 bp downstream from the splice acceptor site found in all C subtype sequences from this study, C2220 and BR025. This clone is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p93IN905.

C_IN.21068 AF067155 India Lole, K.S. *J Virol* **73**(1):152–60 (1999)

This subtype C sequence is one of several complete genomes from India. It is derived from primary PBMC coculture taken Feb. 18, 1995 from a 21 year old man from Pune in Maharashtra State, India who seroconverted in 1994. His only identified risk factor for HIV infection was genital ulcer disease. There is a termination codon at 251 bp downstream from the splice acceptor site found in all C subtype sequences from this study, C2220 and BR025. This clone is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p95IN21068.

C_IN.301999 AF067154 India Lole, K.S. *J Virol* **73**(1):152–60 (1999)

This subtype C sequence is one of several complete genomes from India. It is derived from primary PBMC coculture taken April 24, 1993 from a 52 year old man from Pune in Maharashtra State, India who seroconverted in 1992; risk factors included sex with men and commercial sex workers; NSI phenotype. There is a termination codon at 251 bp downstream from the splice acceptor site found in all C subtype sequences from this study, C2220 and BR025. In clone 93IN999 the vpu start codon is replaced by ATA, thought to modulate the relative expression of Vpu and Env from the same spliced mRNA. This clone is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p93IN999.

C_BR.92BR025.8 U52953 Brazil Gao, F. *J Virol* **72**(7), 5680–98 (1998)

This sequence is from a PCR clone from a primary isolate that is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, *ARHR* **10**:1327–1344 (1994)). 92BR025 was from a 23 year old male hemophiliac from Porto Alegre, Brazil. He had recently seroconverted, though sometime more than 1.2 months prior to the date this blood sample was collected in 1992. He was asymptomatic, and had not taken any anti-retroviral therapy prior to sampling. The isolate 92BR025 was established and propagated by short-term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. The HIV isolate BR025 exhibited an NSI phenotype when assayed by the WHO. The clone has two inframe stop codons in pol at positions 2141, and 3115, and a frame shift mutation at position 4131. Additional env, nef, and ltr region sequences are available from this isolate: U09126, U09132, U51282, and U15121.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
C_ET.ETH2220	U46016	Ethiopia	Salminen, M. *ARHR* **12**:1329–1339 (1996)

ETH2220 was the first reported near full length subtype C sequence. The patient sample from which this clone was obtained was taken in 1986, in Ethiopia. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-κB instead of containing two, a feature which was preserved in other Ethiopian C subtype samples, as well as C viruses from Zambia. This sequence was cloned as a PCR amplified near full length genome, and has a defective tat gene.

D_ZR.Z2Z6 M22639 Zaire Theodore, T. and A. Buckler-White, Unpublished (1988)

An infectious molecular clone of this virus was created by reconstruction. It was cloned in phage and sequenced from the isolate Z2, also called CDC-Z34. All reading frames in this sequence are intact. The entries K03458 and M16322 are from the same isolate, but are defective in Vpr. The former Zaire is now the Democratic Republic of the Congo.

D_ZR.ELI K03454 Zaire Alizon, M. *Cell* **46**:63–74 (1986)

This sequence is of a phage clone derived from the Zairean isolate ELI. ELI was recovered in 1983 from a 24 year old woman with AIDS. All reading frames are intact. A reconstructed infectious clone is available. Gag (M27954) and env (M27949) sequences from the same isolate are also available. The former Zaire is now the Democratic Republic of the Congo.

D_ZR.NDK M27323 Zaire Spire, B. *Gene* **81**:275–284 (1989)

This is an infectious molecular clone derived from a very cytopathic isolate. It was cloned in phage and is replication competent. All reading frames in this sequence are intact. Spire reported that only minor sequence differences appear to be responsible for the “acute biological effect”. This sequences clusters with HIV-1 subtype D in phylogenetic analysis. The former Zaire is now the Democratic Republic of the Congo.

D_ZR.84ZR085 U88822 Zaire Gao, F. *J Virol* **72**(7), 5680–98 (1998)

Sample 84ZR085 was obtained from an AIDS patient from Zaire in 1984. The near full length genome was cloned in phage and sequenced. D-84ZR085 is subtype D. There was a frame shift mutation in gag/pol, position 1692. This isolate was obtained from Thomas Jefferson University, and isolate phenotyping information was not available.

D_UG.94UG114.1 U88824 Uganda Gao, F. *J Virol* **72**, 5680–98 (1998)

Sample 94UG114 was obtained from an asymptomatic 31-year-old man from Butuku, Uganda, as part of the WHO/UNAIDS study. He had not taken any anti-retroviral therapy prior to sampling. His risk factor for infection was heterosexual contact. The near full length genome was PCR amplified from a short term culture of a PBMC sample and sequenced. 92UG114.1 is subtype D. There were no defective gene products, and an infectious molecular clone of 94UG114.1 is available. The isolate from which this sequence was derived is NSI by an MT-2 assay.

F_BE.VI850 AF077336 Belgium Carr,J.K., Unpublished (1998)

Small sections of the 5′ and 3′ LTRs are not included in this sequence. This sequence was isolated from a Belgian man in 1993 whose wife was infected in Zaire (now called the Democratic Republic of the Congo). This sequence was kindly provided prior to publication by J. Carr

Table 2 (cont.)

Sequence	Accession	Origin	Reference
F_FI.FIN9363	AF075703	Finland	Laukkanen,T. Unpublished (1998)

This virus was isolated in 1993 from a Finnish male who was most likely infected in Finland in 1985 by a Kenyan woman. This sequence was kindly provided prior to publication by M. Salminen

F_BR.93BR020.1 AF005494 Brazil Gao, F. *J Virol* **72**(7), 5680–98 (1998)

This isolate is part of a part of a set obtained through the WHO Global Programme on AIDS (WHO Network, *ARHR* **10**:1327–1344 (1994)) and came from an asymptomatic HIV seropositive 52 year old man from Rio de Janeiro, Brazil, sampled in 1993. The risk factor for infection was bisexual contact. The isolate 92BR020 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. There were no defective genes. The isolate 93BR020 was described as syncytium inducing (SI) using an MT-2 assay. An envelope gene sequence from this isolate is described in Gao, F., *J Virol* **70** 1651–1657 (1996). the full length genome is described in Gao, F. *J Virol* **72**(7), 5680–98 (1998).

G_FI.HH8793.11 AF061640 Finland Carr, J.K., *Virology* **247**(1), 22–31 (1998)

This sequence is from an asymptomatic Finnish woman who had sexual exposure to a HIV-1 seropositive man who had lived in Kenya. The proviral DNA was cloned by extended PCR from total genomic DNA extracted from short-term PBMC co-culture. The four full length G subtype sequences share a complex genetic pattern (DRCBL (AF084936), 2NG083 (U88826), HH8793 (AF061640) and SE6165 (AF061642)) with A or AE clustering genomic regions in gp41 and vpr. The See J. Carr *Virology* **247**(1), 22–31 (1998) and F. Gao *J Virol* **72**(7), 5680–5698 (1998) for analyses of the other genomes with this pattern.

G_SE.SE6165 AF061642 Sweden Carr, J.K., *Virology* **247**(1), 22–31 (1998)

This patient (6165) was infected in Congo and moved to Sweden, where he was the index case for a transmission cluster described in Leitner *Virology* **209**:136–146 (1995). He had tested HIV positive (ELISA and Western Blot) approximately 18 months prior to infecting patient 6167 via heterosexual intercourse and 19 months prior to infecting patient 6168, also via heterosexual intercourse. See accession numbers L40744, L40745, L40753, L40754, L40762 and L40763 for sequences from these patients. He had low (35 per μl) CD4 count and dermatological problems but no AIDS defining illness at the time of sampling in 1993. The four full length G subtype sequences share a complex genetic pattern (DRCBL (AF084936), 2NG083 (U88826), HH8793 (AF061640) and SE6165 (AF061642)) with A or AE clustering genomic regions in gp41 and vpr. See J. Carr *Virology* **247** 22–31 (1998) and F. Gao *J Virol* **72**(7), 5680–5698 (1998) for analyses of the other genomes with this pattern.

G_BE.DRCBL AF084936 Belgium Debyser,Z., *ARHR* **14**(5):453–9 (1998)

Clinical details are discussed in *ARHR* **14**(5):453–9 (1998) and the analysis of the complete genome is in press (1999). A pregnant 26 year old women was sampled who had lived in Zaire (now called the Democratic Republic of the Congo) until 1993, then moved to Belgium. She was diagnosed with AIDS and had a low CD4 when sampled in 1996. Her G subtype virus was not detected by Amplicor Monitor or Nasba RNA kits, although she was found to have a high viral load by branched DNA. The four full length G subtype sequences share a complex genetic pattern (DRCBL (AF084936), 82NG083 (U88826), HH8793 (AF061640) and SE6165 (AF061642)) with A or AE clustering genomic regions in gp41 and vpr. See J. Carr *Virology* **247** 22–31 (1998) and F. Gao *J Virol* **72**(7), 5680–5698 (1998) for analyses of the genomes with this A-G pattern. The sequence was kindly provided prior to publication by R. Oelrichs

Table 2 (cont.)

Sequence	Accession	Origin	Reference
H_BE.VI991	none-yet	Belgium	Laukkanen,T. Unpublished (1998)

This subtype H sequence is from a Belgian sample. It was kindly supplied prior to publication by M. Salminen.

H_BE.VI997	none-yet	Belgium	Laukkanen,T. Unpublished (1998)

This subtype H sequence is from a Belgian sample. It was kindly supplied prior to publication by M. Salminen.

H_CF.90CF056.1	AF005496	Central African Republic	Gao, F. *J Virol* **72**(7), 5680–98 (1998)

This sequence clusters with available HIV-1 subtype H sequences in phylogenetic analysis, and was the first available full length H subtype sequence . The isolate comes from Bangui, in the Central African Republic, and was sampled in 1990, from an asymptomatic individual, who had no anti-retroviral therapy. The isolate had an NSI phenotype by an MT-2 assay, and the sample was obtained from the Pasteur Institute, Bangui. The isolate 90CF056 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified, cloned and sequenced. The isolate was at one point designated 90CR056, but was changed to 90CF056 as CR stands for Costa Rica, and CF for Central African Republic. The first genetic characterization of this virus isolate (an env V3 sequence designated 4056, GenBank accession number L11497, Murphy, E., *ARHR* **9**:997–1006 (1993)) left the subtype designation as unclassified, but a second study of this env region sequence classified it as subtype H (Janssens, W., *ARHR* **10**:877–879 (1994)). An additional ltr (U51290) and env and nef genes are available from this isolate (U08797 and U27401). The Gao 1998 publication of this sequence provided the first full length genome of this subtype. There were no defective genes in the sequence.

J_SE.SE91733	AF082395	Sweden	Laukkanen,T., *ARHR* In Press (1998)

This sequence is from a woman who was infected in Zaire between 1981 and 1986. Blood for sequencing was drawn in 1993. She was asymptomatic with a CD4 count of 184. The sequence was kindly provided prior to publication by M. Salminen. This sequence is from the same individual as SE7022 described by T. Leitner *ARHR* **11**(8):995–997 (1985), see accession numbers L41177 and L41179 for env and gag genes from this individual. Other examples of subtype J have been found in Gambia, see accession numbers U33099, U33100 and U33102.

J_SE.SE92809	AF082394	Sweden	Laukkanen,T., *ARHR* In Press (1998)

This sequence is from a male who was infected in Sweden between 1993 and 1994. Blood for sequencing was drawn in 1994. He was asymptomatic with a CD4 count of 567. The sequence was kindly provided prior to publication by M. Salminen. This sequence is from the same individual as SE7887 described by T. Leitner *ARHR* **11**(8):995–997 (1985), see accession numbers L41176 and L41178 for env and gag genes from this individual.

Inter-subtype Recombinants

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AB_RU.KAL153–2	none-yet	Russia	Salminen, M., Unpublished (1998)

This is the first full length sequence of the subtype A/B recombinant strain that is the circulating recombinant form responsible for an explosive epidemic among injecting drug users in Kaliningrad, Russia (Liitsola K *AIDS* **12**(14):1907–19 (1998)). Newly diagnosed HIV-1+ individuals in Kaliningrad rose from less than one per month to more than 100 per month during the period of July-October 1996, with a 1335 new infections identified between 1 July 1996 and 30 June 1997. The subtype A region of the virus was similar to A subtype sequences found in IDUs in the Ukraine, and the subtype B region is of unknown origin.

AC_ZM.ZAM184 U86780 Zambia Salminen, M., *J Virol* **71**:2647–2655 (1997)

The envelope gene from this sample taken from a Zambian woman in 1990 (Louwagie, J., *J Virol* **69**:263–271, 1995) proved to be an A/C recombinant in subsequent phylogenetic analyses (D. Robertson *Nature* **374**:124–126 (1995)). The full length provirus was recovered using PCR, and sequenced, according to the method described in Salminen, M., *Virology* **213**:80–86 (1995). Additional clones from serial samples from the index case ZAM184, and from her spouse who was also HIV-1 positive (accession numbers U86768-U86781 and L22955) represent gag, env, or additional full sequences from this couple. The A/C mosaic pattern of the full length genomic sequence is presented in Salminen, M., *J Virol* **71**:2647–2655 (1997), and also in Robertson, D., part III pages 25–30, of the 1997 compendium. All coding sequences are intact.

AC_RW92RW009.6 U88823 Rwanda Gao, F., *J Virol* **72**(7), 5680–98 (1998)

The isolate 92RW009 was part of the set generated through the WHO Global Programme on AIDS (see: WHO Network, *ARHR* **10**:1327–1343 (1994)). The virus was derived from an asymptomatic 24 year old woman from Kigali, Rwanda, whose route of infection is thought to be heterosexual contact, and who had no anti-retroviral therapy. The clone has a frameshift mutation in gag at position 213. The blood sample was taken in 1992. The original env and gag sequence from this isolate clustered with HIV-1 subtype A sequences (Gao, F., *ARHR* **10**:1359–1368 (1994)), however subsequent in-depth analysis of the full length genome sequence from this isolate suggests it is an AC mosaic sequence with multiple crossover points. The breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium. The primary isolate from this patient and early passage cultured virus were NSI on MT-2 cells when analyzed by two different WHO labs in 1993–1994 (WHO Network, *ARHR* **10**:1327–1343 (1994), and De Wolf, F., *ARHR* **10**:1387–1400 (1994). The NIAID 1997 Reference Reagent Catalog classifies it as NSI. However, more recent papers (Zhang L., *Nature* **383**:768 (1996) and Zhang L., *ARHR* 13(16), 1357–1366 (1997), classify it as SI. This sequence is from PCR-amplified proviral DNA harvested from PBMCs from the NIH NIAID reagent repository. The NSI phenotyping was determined using earlier passages, and SI was associated with later passages, (J. Bradac, personal communication). Both the full length clone 92RW009.6 and the partial gp120 clone 92RW009.14 (accession U08793) were amplified from the same DNA extracted from a short term primary PBMC culture prepared by Ogden Bioservices, stored in the NIAID Research Reagent Repository, lot 1. Other env sequences from this isolate are: U16221, U08793, U13441, U16220, and U16222, and a gag sequence is U86545.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AC_IN.IN2103	AF067156	India	Lole, K.S. *J Virol* **73**(1):152–60 (1999)

A small section of the 5′ LTR sequence present in the 21301 virus is not included in this sequence. This subtype AC sequence is one of several complete genomes from India. It is derived from primary PBMC cocultures taken April 3, 1993 from a 40 year old man from Pune in Maharashtra State, India who seroconverted in 1995; his risk factor was sex with commercial sex workers. This sample is available in plasmid form through the NIH AIDS Research and Reference Reagent Program and is named p95IN21301.

ADI_ZR.MAL K03456 Zaire Alizon, M. *Cell* **46**:63–74 (1986)

This sequence is from a lambda phage clone derived from the Zairean isolate MAL. MAL was recovered in 1985 from a 7 year old boy with ARC, probably infected by a blood transfusion in 1981, as his parents were seronegative. All reading frames are intact except for vpu, due to the loss of a start codon. A reconstructed infectious clone is available. MAL was one of the first African sequences characterized, and soon after the initial characterization it was determined to be a mosaic (Li, W.-H., *Genetics* **116**:s44 (1987)). Recent analysis suggests that it has three different distinct subtype associations, A, D and I, and some regions that are difficult to characterize as associated with any known subtype (see Robertson, D., part III pages 25–30, of the 1997 compendium). The I association is based on phylogenetic associations with a clone from Cyprus isolate that was the first I subtype characterized (94CY032). The 94CY032 clone also appears to be an A-G-I mosaic (Robertson, D., part III pages 25–30, of the 1997 compendium, and Gao, F., *J Virol* **72**(12):10234–41 (1998).

AE_CF.90CF402.1 U51188 Central African Republic Gao, F. *J Virol* **70**:7013–7029 (1996)

The isolate 90CF402 was sampled from a man from Bangui, the Central African Republic, who was suffering from AIDS-related conditions. His risk factor for infection was heterosexual contact. The isolate was first adapted to growth in chimpanzee cells, then re-expanded in human PBMC before a lambda phage clone was generated and sequenced. The sequence has a defective vpu due to the lack of a start codon, and a defective vif gene. A reconstructed infectious molecular clone is available. The pattern of subtype A-E recombination breakpoints is shared between A-E subtype sequences from Thailand and from the Central African Republic, suggesting a shared ancestral recombined virus that arose prior to the subsequent epidemics in the two areas. The AE breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, Gao, F. *J Virol* **70**:7013–7029 (1996), and Carr, J. *J Virol* **70**:5935–5943 (1996). The prototype of this circulating recombinant is the sequence CM240, and sequences of with this genetic pattern have historically been called subtype E.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AE_TH.CM240	U54771	Thailand	Carr, J. *J Virol* **70**:5935–5943 (1996)

Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. The blood sample was drawn in 1993. DNA from p24 antigen-positive culture was used to PCR amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and E virus, with E clade dominating in env and the LTR, and A in in gag and pol. This is the pattern of A-E sequences found throughout Asia and Africa, and no full length E subtype reference strain has been identified. Carr provide detailed analysis of the breakpoints, and point out that the A/E mosaic genomes have a natural pseudotype structure where the external envelope protein spikes on the virion essentially are contributed by the E subtype, and the rest of the viral proteins have a subtype A origin. The AE breakpoints are also mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, and Gao, F. *J Virol* **70**:7013–7029 (1996). See also the env sequence from the same isolate (L14572), Mascola, J., *JID* **169**:48–54 (1993). This sequence is the prototype sequence of the AE circulating recombinant form that is found in Asia and central Africa, and sequences of with this genetic pattern have historically been called subtype E.

AE_TH.93TH253.3 U51189 Thailand Gao, F. *J Virol* **70**:7013–7029 (1996)

This virus was isolated in 1993 from a 21 year old man from Chiang Mai, Thailand, who had end-stage AIDS. The isolate was previously designated CMU010, or 302053. The isolate is NSI. It was expanded in donor PBMCs, then in H9 cells, then a lambda phage clone was generated and sequenced. The sequence has a defective vpu due to the lack of a start codon, and a defective env gene. Like other "E" subtype viruses from both Asia and Africa, large stretches of the genome are associated with the A subtype, and all share a common mosaic pattern of A/E breakpoints, suggesting that the currently identified A-E recombinants all share a common ancestor. The AE breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, Gao, F. *J Virol* **70**:7013–7029 (1996), Carr, J. *J Virol* **70**:5935–5943 (1996). The prototype of this circulating recombinant form is the sequence CM240, and sequences of with this genetic pattern have historically been called subtype E.

AG_NG.IBNG L39106 Nigeria Howard, T. *ARHR* **12**:1413–1425 (1996)

HIV-1 sample IbNg was isolated from the PBMCs of an asymptomatic 23 year old man from Nigeria. DNA from this isolate was PCR amplified and cloned, with the complete genome sequence derived from multiple PCR amplification products. The partial env gene sequence (U48628) was originally designated subtype A (Howard, T. *ARHR* **10**:1755–1757 (1994)), as was the full length genome. The full length sequence was eventually shown to be an A-G recombinant with multiple cross-over points. The breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, and Gao F, *J Virol* **70**:7013 (1996). The IbNg sequence has a 16 bp insertion within the Lys-tRNA primer binding site, just 3′ of the 5′ LTR. It also has a single nucleotide deletion in tat cds at position 5449. There are several sequences which share recombination breakpoints with IbNG, and are essentially the same recombinant circulating form; IbNG is the prototype, sharing a similar structure with DJ263 and DJ264 (GB Accession numbers AF063223 and AF063224 Carr *Virology* **247**:22–31 (1998)).

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AG_FR.DJ264	AF063224	Djbouti	Carr *Virology* **247**:22–31 (1998)

A small section of LTR sequence present in the DJ263 virus is not included in this file. Carr states that this virus was from a French foreign legion soldier assigned to peacekeeping duties in Djbouti, referencing Louwagie *J Virol* **69**(1), 263–271 (1995). However the Louwagie paper does not mention the French soldiers and only states that the blood sample was from Djbouti. The sample was isolated in 1991. There are several sequences which share AG recombination breakpoints with IbNG, and are essentially the same recombinant recirculating form; IbNG is the prototype, sharing a similar structure with DJ264 and DJ263. Also see L23064

AG_FR.DJ263 AF063223 Djbouti Carr *Virology* **247**:22–31 (1998)

Carr states that this virus was from a French foreign legion soldier assigned to peacekeeping duties in Djbouti, referencing Louwagie *J Virol* **69**(1), 263–271 (1995). However the Louwagie paper does not mention the French soldiers and only states that the blood sample was from Djbouti. The sample was isolated in 1991. There are several sequences which share AG recombination breakpoints with IbNG, and are essentially the same recombinant recirculating form; IbNG is the prototype, sharing a similar structure with DJ264 and DJ263. Also see L22941

AG_NG.92NG003.1 U88825 Nigeria Gao, F. *J Virol* **72**(7), 5680–98 (1998)

This sequence is from a PCR clone from an NSI primary culture from isolate G3, renamed 92NG003 to be consistent with WHO nomenclature. The sample was taken in 1992 from a 27 year old, asymptomatic HIV seropositive female prostitute from Jos, Nigeria. (Abimiku, A., *ARHR* **10**:1581–1583 (1994), env sequence accession number U13208). The isolate came from the Institute of Human Virology, Baltimore, MD, and is NSI. While originally described as subtype G in env, this genome has mosaic A/G regions, with A like regions in the central part of the genome. The breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, and in Gao 1998. There are frameshift mutations associated with 10–16 bp deletions in vpr and vpu, at positions 5024 and 5485, as well as deletions totaling 33 bp near the 3′ end of the V3 loop. Nef has an altered initiation codon at position 8113.

AGI_ZR.Z321 U76035 Zaire Choi, D. *ARHR* **13**:357–361 (1997)

Z321B is a later passage of isolate Z321 (see GenBank Accession Numbers M15896, U50208, and U50207); Z321 was isolated from a 1976 Zairean serum sample. Z321 was grown to industrial scale as a chronically infected T-cell line to manufacture an inactive, therapeutic HIV-1 immunogen. Z321B was then established from this industrial scale stock. The genomic sequence was derived from multiple PCR clones amplified from Z321B. Z321B contains a mutation in the termination codon of the tat gene, (bases 2294–2296) so that the tat gene of HIVZ321 extends further downstream (bases 2342–2344), and has a defective vpr and vpu. Recent analysis suggests that it has three different distinct subtype associations, A, G and I, and some regions that are difficult to characterize as part of any known subtype (see Robertson, D., part III page 25–30 of the 1997 compendium). The I association is based on phylogenetic associations with a clone from Cyprus isolate that was the first I subtype characterized (94CY032). The 94CY032 clone also appears to be an A-G-I mosaic (Robertson, D., part III pages 25–30, of the 1997 compendium) but with different recombination breakpoints, and Gao, F., *J Virol* **72**(12):10234–41 (1998).

Table 2 (cont.)

Sequence	Accession	Origin	Reference

AGI_CY.94CY0323 AF049337 Gao, F., *J Virol* **72**(12):10234–41 (1998)

This sample, like others in this study (see also subtypes A, B, C, and F) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. Patient HO31 was a 24 year old asymptomatic female known to have been HIV seropositive for at least 5 years. Patient HO32 was a 35 year old asymptomatic male, also seropositive for at least 5 years. Both were IVDUs who had lived in Greece and used IV drugs there, before moving to Cyprus. DNA was extracted from patient PBMCs and PCR amplified. Products were cloned and sequenced. Two env gene clones from HO32 and one from HO31 were sequenced (accession numbers U28672, U28673 and U28685). Complete genome is only available for HO32 (CY032). For patient information see Kostrikis,L.G. *J Virol* **69**:6122–6130 (1995). This sequence has the same genetic recombination breakpoints as PVCH and PVMY, and CM240 is the prototype of the circulating recombinant form. The Gao 1998 paper characterizes subtype I in greater detail, presenting the fist published account of this full length genome. The analysis of C2-V3 env gene sequences confirmed that 94CY032.3 was closely related to sequences previously classified as subtype I. However, the remainder of its genome various regions in which 94CY032.3 was significantly clustered with either subtype A or subtype G. Only regions in vpr, nef, and the middle portions of pol and env, formed independent lineages roughly equidistant from all other known subtypes. Since these latter regions most likely have a common origin, Gao classified them all as subtype I, and report that 94CY032 represents a triple recombinant (A/G/I) with at least 11 points of recombination crossover.

AGI_GR.97PVCH AF119820 Nasioulas, G. *ARHR* (in press) (1998)

G. Nasioulas, D. Paraskevis, E. Magiorkinis, M. Theodoridou and A. Hatzakis “Molecular Analysis of the full-length Genome of the HIV-1 Subtype I: Evidence of Triple Recombination” (submitted for publication) (1998) The sequence isolated from patient GR11 (accession AF049292) is from the same patient as 97PVCH AF049292. The patient was a 32 year old male IVDU with symptoms (CDC stage B3) in 1991, when sampled. He is no longer living. For patient information see information on patient GR11 in Nasioulas, G.*ARHR* **14**(8), 685–90 (1998). This sequence has the same genetic recombination breakpoints as CY032 and PVMY, and is one of the AGI circulating recombinant forms of which CY032 is the prototype.

AGI_GR.97PVMY AF119819 Nasioulas, G. *ARHR* (in press) (1998)

G. Nasioulas, D. Paraskevis, E. Magiorkinis, M. Theodoridou and A. Hatzakis “Molecular Analysis of the full-length Genome of the HIV-1 Subtype I: Evidence of Triple Recombination” (1998). The sequence was isolated from a 13 year old whose mother and father were IVDUs. The isolate is also called GR84. This sequence has the same genetic recombination breakpoints as CY032 and PVCH, and is one of the AGI circulating recombinant forms of which CY032 is the prototype.

AGJ_BF.BFP90 AF064699 Oelrichs, R.B. *ARHR* (in press) (1998)

HIV-1 from Burkina Faso, identified in Australia. The patient is a 32 year old African male who acquired the infection heterosexually in 1991. The patient was diagnosed in August 1996 at which time he had a CD4 count of 125. This sequence was derived by PCR directly from patient PBMCs when therapy-naive. The genome of this virus is a mosaic between subtypes A, G and J. The LTR is subtype J, a segment of about 950 bp at the beginning of gag is subtype A, the remainder of gag and part of the protease sequence are subtype G. Most of the pol gene cannot be assigned to a subtype. The mid-genome accessory region is mostly subtype J. gp120 is subtype G. The 3′ region of gp41, the third exons of tat and rev, and the nef gene are subtype J.

Table 2 (cont.)

Sequence	Accession	Origin	Reference
BF_BR.93BR029.4	AF005495	Brazil	Gao, F. *J Virol* **72**(7), 5680–98 (1998)

The isolate 93BR029 is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, *ARHR* **10**:1327–1344 (1994)), and came from an asymptomatic HIV seropositive 17 year old male, with unknown risk factor, from Sao Paulo, Brazil, sampled in 1993. The isolate had an NSI phenotype by an MT-2 assay. The isolate 93BR029 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. An envelope gene sequence from this isolate is described in Gao, F., *J Virol* **70**, 1651–1657 (1996), U27413, and an LTR sequence is also available U51291. The envelope gene was first described as an F subtype, however subsequent phylogenetic analysis of the full length genome indicated that the clone is a mosaic with regions of B and F subtype. The breakpoints are mapped in Robertson, D., part III pages 25–30, of the 1997 compendium, and in Gao 1998. There are two frame shift mutations in gag, in positions 269 and 472.

N_CM.YBF30 AJ006022 Simon, F. *Nature Medicine* **4**(9), 1032–1037 (1998)

This genome is from a 40 year old Cameroonian woman whose blood was sampled in May and December, 1995. She died in December of 1995 with AIDS. YBF30 was derived from the May sample. The isolate was generated after coculture of with PHA-stimulated donor PBMCs. YBF30 used the CCR5 coreceptor and did not use CXCR4. Of 700 Cameroonian blood samples analyzed in this study, 16 (2%) were more reactive with SIV-CPZ peptides than with HIV-1 group M or group O peptides. Of those 16, 3 were strongly reactive with YBF30 peptides. One of them, YBF105, was sequenced in Pol and confirmed to belong to the N group.

O_CM.MVP5180 L20571 Cameroon Gurtler, L. *J Virol* **68**:1581–1585 (1994)

The isolate MVP5180 was derived from a Cameroonian woman who had AIDS in 1991; she died of AIDS in 1992. The isolate from which the clone was derived was grown in several T cell lines, and could also grow in the monocytic U937 cells.

O_CM.ANT70 L20587 Cameroon Vanden Haesevelde, M. *J Virol* **68**:1586–1596 (1994)

ANT70 was isolated from the first O group infection discovered, and the very divergent LTR sequence was first published in 1990 (de Leys, R., *J Virol* **64**:1207–1216, 1990). The isolate came from a CDC stage III infected individual with unusual serological reactivity. O group viruses have the same genetic organization as M group viruses, which dominate the epidemic, but are quite distinct in terms of their genetic sequences. For a review, See Korber B., Human Retroviruses and AIDS Database, Part III, 41–56, 1996

SIV-CPZANT U42720 Zaire Vanden Haesevelde, M. *Virology* **221**:346–350 (1996)

The CPZANT was isolated from a wild caught chimpanzee from Zaire. Two additional SIV CPZ (chimp) viral isolates have been generated from chimps caught in Gabon (U11495, X52154). The chimpanzee viral sequences are genetically more closely related to the HIV-1 sequences derived from infected humans than are HIV-2 strains or other SIVs.

SIV-CPZGAB X52154 Gabon Huet, T. *Nature* **345**:356–359 (1990)

The CPZGAB virus was isolated from a chimpanzee caught in Gabon. The genome is more closely related to HIV-1 than are HIV-2 or other SIV viral sequences. Also see CPZGAB2, U11495 for a sequence fragment from an additional chimpanzee caught in Gabon.