

A Subtype

At this time there are viral sequences from 539 HIV-1 infected individuals associated with HIV-1 subtype A. The A subtype consensus sequence (A_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BE1.ID#:** These 29 sequences are from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. One of these patients (VI822) was triple infected with subtypes A, B and E. Two of the patients (VI712 and VI832) had nearly identical sequences, so only VI712 is presented here. Several of these sequences cluster with the CRF02(AG) sequences in phylogenetic analysis, for example VI712. Accession numbers AJ228214, AJ228194, AJ228150, AJ228152, AJ228180, AJ228201, AJ228217, AJ228226, AJ228156, AJ228235, AJ228207, AJ228233, AJ228195, AJ228211, AJ228167, AJ228186, AJ228234, AJ228209, AJ228218, AJ228176, AJ228231, AJ228159, AJ228220, AJ228205, AJ228157, AJ228206, AJ228215, AJ228221, AJ228232 and AJ228224.
- 2) **BJ1.ID#:** These 18 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. 15 are from directly sequenced PCR products, derived via RT-PCR from patient serum RNA. 3 (233, 251 and 253) are from cloned PCR products, also by RT-PCR from serum RNA. Another subtype A sequence (366, U61870) is not included here, because it was nearly identical (254 of 255 bases identical) to the SF170 (M66533) sequence from Rwanda isolated in 1988, and thus it likely represents a lab artifact. [Heyndrickx (1996)]. Accession numbers U61854–U61869, U61871 and U61873. Two subtype G isolates were also found in this study.
- 3) **BY.BLR10A:** This sequence is from Byelorussia [Lukashov (1995)]. Accession number L38411.
- 4) **CA.HWCL1:** This sequence is the first published sequence of subtype A HIV-1 in Canada. The patient had moved from Uganda in 1983, and was diagnosed as HIV+ in 1989. Viral RNA was recovered from archived, stored patient serum by binding viral particles to CD4-coated wells of an ELISA plate. The RT-PCR amplification product was cloned and 10 clones were sequenced. It is not clear whether this sequence is from a single clone, or a consensus of all 10 [Montpetit(1995)]. Accession number U34049.
- 5) **CD1.ID#:** These nine sequences are part of a set of 14 A and D sequences from women from Democratic Republic of Congo (formerly Zaire); 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. PCR-direct, peripheral blood DNA. [Potts (1993a)]. L19633 was too short to include here, but was also subtype A. Accession numbers L19624–L19626, L19628–L19630, L19632–L19634, and L19636.
- 6) **CD2.ID#:** These 4 sequences are from Democratic Republic of Congo (formerly Zaire) (Reitz et al, unpublished 1993). Accession numbers U43097–U43100.
- 7) **CD3.ID#:** These 13 subtype A sequences are from samples collected in the Democratic Republic of the Congo (Takehisa et al unpublished 1999). Accession numbers for subtype A are AF119189, AF119190, AF119193, AF119195, AF119200, AF119201, AF119202, AF119204, AF119205, AF119206, AF119207 and AF119209.
- 8) **CF1.ID#:** These thirteen sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. The sequences are from cloned PCR products. DNA was isolated from co-cultured PBMCs [Murphy (1993)] Accession numbers L11457–L11458, L11461–L11462, L11469–L11471, L11474–L11475, L11477–L11479, L11484–L11496, L11498, L11518 and L11523–L11524. Some of them were sequenced again by Schmitt et al. (unpublished): U43275, CF1.4023; U43109, CF1.4054; U43139, CF1.11423; U43136, CF1.1286. Some of these sequences were re-analyzed along with more recent Central African Republic sequences in [Muller-Trutwin (1999)].
- 9) **CF2.ID#:** These 2 sequences were kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. Accession numbers U43111, U43171, M81063, M81064.

- 10) **CF3.ID#:** These 25 sequences are from a set of sequences obtained from patients from the Central African Republic [Muller-Trutwin (1999)]. Accession numbers AF067735–AF067751, AF067756–AF067758 and AF067762–AF067765.
- 11) **CG1.ID#:** These 7 sequences are from 1988 and 1992 samples collected in Brazzaville and Pointe Noire, Congo [Candotti (1991), Candotti (1999)]. All patients were adults with AIDS at the time of sampling. Subtypes A, D, F, G and recombinants (CRF01 and CRF02) were identified in this study. Accession numbers for the subtype A samples are AF082307–AF082308, AF082310, AF082312, AF082314–AF082315 and AF082319. Gag sequences are also available for some of the samples, with accession numbers M73472–M73480.
- 12) **CI1.ID#:** These six sequences are from six different AIDS patients suffering from pulmonary tuberculosis at the Pneumology Hospital of Cocody, Abidjan, Ivory Coast. In this study sequences came from coculture (on donor PBMCs) supernatant viral RNA, proviral DNA from each patients' PBMCs after coculture with donor PBMCs, and proviral DNA directly from uncultured PBMCs. PCR or RT-PCR was used to amplify the env V3 region, and 4-7 cloned PCR products were sequenced. A total of 66 sequences from the six patients were published. All 66 were subtype A, and inpatient sequences were more similar than interpatient sequences [Audoly (1996)]. Accession numbers U59559–U59624.
- 13) **CI2.ID#:** These 3 sequences are from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-14 and CI-20 were symptomatic, and the others were asymptomatic. CI-14, CI-45 and CI-47 were serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp sequence. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the one of those clones is presented here. [Janssens (1994a)]. Several of these sequences which were previously classified as subtype A, have been reclassified as CRF02(AG) based on more recent phylogenetic analyses. Accession numbers X72037–X72039, X72043–X72056, X72052–X72054 and X72062–X72065.
- 14) **CI3.ID#:** These 21 sequences are from Abidjan, Ivory Coast. Subtypes A, D and G were found for Ivory Coast patients in this set [Ellenberger (1999)]. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000449, AF000450, AF000452, AF000453, AF000454–AF000457, AF000459, AF000461, AF000462, AF000464–AF000467, AF000470–AF000475 and AF000493.
- 15) **CM.CMR61:** CMR61 is one of two A group sequences from a 23 year old female sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa (1997b)]. The other subtype A sequence from the same patient is labelled as CMR709 and is from a separate blood sample. Accession numbers U58148, U58150. Entries labeled as CMR709 (AF055728, AF055730) and CMR1000 (AF097692, AF097697) are 97% identical to CMR61.
- 16) **CM1.ID#:** These sequences are 10 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic (CA7, CA11, CA15, CA17, CA18, and CA21) and symptomatic (CA2, CA6, CA19 and CA22) individuals. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate [Nkengasong (1994)]. The other six sequences were subtypes B, E, F, H and AGU-recombinant. Accession numbers for the subtype A sequences: X80440, X80442, X80444–X80447, X80449, X80450, X80453, X80454.
- 17) **CM3.ID#:** These 15 sequences are all from Cameroon (Takehisa et al, unpublished 1997). Accession numbers U69992–U69996, U70000, U70002–U70008, U70010 and U70014. Several of these sequences, for example U70002 and U70008 are most likely CRF02(AG) circulating recombinant forms.
- 18) **CM4.ID#:** These 17 subtype A sequences are from 1994–1995 samples from 211 Cameroonian AIDS patients [Takehisa (1998)]. Of the 43 HIV patients sequenced, 17 were subtype A, 1 was subtype B, 2 were subtype C, 1 was subtype G. Several of these subtype A sequences cluster with the CRF02(AG) IBNG-like sequences in phylogenetic analysis, but they are left here in subtype A, because it is not clear if they are truly AG recombinant. Accession numbers for subtype A are AF023064–AF023071 and AF023073–AF023081.
- 19) **CM5.ID#:** These 5 subtype A sequences are from samples collected in 1995 from 53 northern Cameroonian Muslim AIDS patients [Mboudjeka (1999)]. Of the 16 HIV patients sequenced, 5 were subtype A, 6 were CRF02(AG) IbNG-like, 2 were subtype D, 1 was subtype G and 2 were subtype H. Accession numbers for subtype A are AF028319, AF028324, AF028327, AF028328 and AF028332.

- 20) **CU1.ID#:** This sequence is from Cuba (Gomez unpublished 1997). Accession number Y14412. The other subtype A sequence in this study (93CU044, Y14418) was more than 99% identical to 92RW002 (U08794) and is therefore not included here.
- 21) **ET.TP95001:** This study describes the distribution of HIV-1 subtypes in Ethiopia. HIV-1 RNA was collected from sera (from a majority of asymptomatic individuals) and RT-PCR amplified and sequenced directly. One subtype A and numerous subtype C sequences were found [Abebe (1997)]. Accession number U88756.
- 22) **FR.DIALM2:** This sequence is from France (Roques et al, unpublished 1998). Accession number AJ006747.
- 23) **FR1.ID#:** These 8 sequences are from members of the French military who are believed to have been infected while deployed outside of France (Djibouti, Guyana, and Central African Republic). Other sequences from this study were subtypes B, C, E, F, and unclassified [Lasky (1997)]. Accession numbers for subtype A were U58800–U58806 and U58778.
- 24) **FR2.ID#:** These 3 subtype A sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers Z95441, Z95442, Z95447 and Z95448.
- 25) **FR3.RO:** This subtype A sequence is from a mother-infant transmission case in France [Pasquier (1998)]. Child RO was born via vaginal delivery and survived more than one year. Three other mother-child pairs were infected with subtype B virus. Accession numbers AJ008753–AJ008773, AJ008835–AJ008856, AJ008963–AJ008985 and AJ009051–AJ009071.
- 26) **GA1.ID#:** These 4 sequences are from Gabon. Two (G41, G135) are from 1988–1989 samples from patients with AIDS living in Franceville Gabon. VI685 is from a 1992 sample of an AIDS patient from the Libreville General Hospital. VI1076 is from a 1993 sample of an AIDS patient from the Libreville General Hospital. LBV23 is from a 1988 sample from an asymptomatic individual sampled from the general population. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Delaporte (1996)]. Accession numbers X90914, LBV23; X90915, G135; X90917, G41; X90924, VI685. See also subtypes C, D, F, A/G, CRF02_AG and O sequences from this same study.
- 27) **GB1.ID#:** These 2 sequences are from British isolates [Douglas (1997)]. Accession numbers Y13717, Y13718.
- 28) **GH2.ID#:** These 8 sequences are from Ghana. Subtypes D, G and an A/G recombinant were also detected in this study [Takehisa (1997a)]. Pol gene sequence is available for GH9 (U67040). Accession numbers for subtype A are U67051–U67054, U75457–U75459.
- 29) **GH3.ID#:** These 13 sequences are from Ghana. Subtypes D and G were also detected in this study [Ishikawa (1996)]. Two additional subtype A sequences are not included here (GH14 U67689 and GH17 U67690) because they were greater than 98% identical to sequences reported by Takehisa et al [Takehisa (1997a)] and may be from the same patients, although labelled differently in the two studies (GH14 here 99% identical to GH15 by Takehisa; GH17 here 98.2% identical to GH16 by Takehisa). Accession numbers for subtype A are U67685–U67700.
- 30) **GH4.ID#:** These 24 sequences are from Ghana. Subtype G was also detected in this study [Bobkov (1998b)]. Many of these sequences seem to be related to the CRF02_AG (IbNG) circulating recombinant form, which is subtype A in the V3 region. But because no gag or pol sequences were determined, these sequences will be considered to be subtype A, rather than A/G recombinant, until more sequence data is available. Accession numbers for subtype A are AJ225657–AJ225681.
- 31) **GH5.ID#:** These 21 sequences are from Ghana [Sagoe (1999)]. Many of these sequences seem to be related to the CRF02_AG (IbNG) circulating recombinant form, which is subtype A in the V3 region. But because no gag or pol sequences were determined, these sequences will be considered to be subtype A, rather than A/G recombinant, until more sequence data is available. Accession numbers AF116875–AF116895.
- 32) **GR.GR48:** This sequence is from northern Greece. The sequence contained numerous frameshifting insertions, possibly due to sequencing errors, [Papa (1998)]. Accession number AJ224956.

- 33) **GR2.ID#**: These 5 sequences are from Greece [Nasioulas (1998)]. Accession numbers AF049296, AF049297, AF049320, AF049321, AF049322.
- 34) **IL.BC41**: This subtype A sequence is from Israel [Gehring (1997)]. Other subtypes found in Israel in this study were B, C, D, K and CRF02(AG). Accession number for the subtype A sequence is X94397.
- 35) **IN1.ID#**: These two entries are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive homosexual men. Other samples from the same region were all subtype C (see C_IN4.#) [Cassol (1996)]. Accession numbers U53286 and U53291.
- 36) **IN2.GT6**: This sequence is from India. Seven other sequences from this publication were subtype C (see C_IN5.ID#) [Tsuchie (1993)]. Accession number D13425.
- 37) **IN3.PTA**: This sequence is from India. Seven other sequences from this publication were subtype C (see C_IN6.ID#), and two were of the B-prime Asian subclade of subtype B. [Maitra (1999)]. The sequence of this subtype A isolate is not included in the alignment, because it is greater than 98% identical to the A_RW.SF170 lab strain of subtype A (M66533), and likely represents a lab artifact. Accession numbers AF101116 and AF101124.
- 38) **IN4.46152**: This sequence is from India. One other subtype A sequence from this set was not included here because it was very similar to this one (46149, AF148250) (Seth, P. et al, unpublished 1999). Other sequences in this set were subtypes B and C (see B_IN3 and C_IN7). Accession number AF148249.
- 39) **KE.K89**: This sequence is named “KENYA” in the GenBank entry, but is identified as K89 in the original manuscript. It is a Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession number L22943. A gag gene sequence from the same isolate is found with accession number L11774.
- 40) **KE1.ID#**: These 6 sequences were derived from patients who were part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from each patient were sequenced. Two other patients from this study had viral subtypes C and D. [Zachar (1996a)]. Accession numbers U32658, U33763, U33764, U33766, U33767 and U34905.
- 41) **KE2.ID#**: These 4 sequences are part of a set of 13 subtype A sequences are from Kenya. The other 9 subtype A sequences were too short to include here. Another 5 sequences from this set were found to be subtype D. The Q23 sequence is from a female recently infected with HIV-1, from Mombasa, Kenya. Blood was drawn on June 13, 1994 for the env region sequences [Poss (1997)] and in July 1993 for a sequence of the complete genome of Q23 (Poss unpublished 1998). Accession numbers AF004885-AF004891, AF03159-AF03161, AF004892, AF004895, AF004897-AF004899. Out of these 14 (13 A and 5 D), 12 were too short to be included (AF004885-AF004891).
- 42) **KE3.ID#**: Although there were no subtype A sequences submitted to the databases from this study, they found that the majority of HIV-1 found in Kenyan women were of subtype A. The women were patients who were part of a study of breastfeeding women from Nairobi, Kenya. Viral DNA was amplified from uncultured patient PBMC, and the envelope V1-V5 region was sequenced after cloning into M13 phage. Patients from this study had viral subtypes A 70%, C 7%, D 20%, G 0.3% and various recombinant forms 2.2% [Neilson (1999)].
- 43) **KE4.ID#**: These 20 sequences were derived from patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced [Janssens (1994c)]. K976 (U12992) was listed as subtype unclassified in the paper, but seems to be subtype A in our analysis. Accession numbers for the subtype A sequences in this study: U12987–U13006.
- 44) **KE5.ID#**: These 21 subtype A sequences are from Kenya [Robbins (1999)]. The samples were collected in 1994 and 1995 from male truck drivers and female sex workers near Mombasa and Nairobi. Subtypes C and D were also found in this study. Accession numbers for the subtype A sequences in this study: AF103910, AF103911, AF103913–AF103915, AF103917, AF103918, AF103921–AF103924, AF103926, AF103928–AF103931, AF103934–AF103938.

- 45) **KR.ROK39:** This sequence is from South Korea [de Souza (1997)]. Like other non subtype B Korean sequences, this one is from a patient believed to have been infected while travelling outside Korea. The patient was a professional sailor who reported having heterosexual contacts in Africa. Accession number U60016.
- 46) **KR1.ID#:** These 3 sequences are from South Korea. Subtypes B, C and H were also reported, but the subtype H sequence (KR68) and one of the subtype A sequences (KR61) were not submitted to the databases [Kim (1999)]. Accession numbers for the entire set are Z92548–Z92668. Subtype A accession numbers are Z92584, Z92607, Z92608, Z92609, Z92645, Z92646.
- 47) **LB1.ID#:** These 10 subtype A sequences are from a study of 25 sequences from a study of multiple HIV-1 subtypes in Lebanon [Pieniazek (1998)]. Of the 26 samples, 25 were classified as HIV-1: 10 were subtype B, 10 were subtype A, 1 was subtype C, 1 was subtype D and 3 were recombinants or untyped. The other sample was classified as HIV-2 subtype B. Accession numbers AF025691, AF025693, AF025695, AF025699, AF025700, AF025703, AF025709, AF025710, AF025711 and AF025712 are HIV-1 subtype A.
- 48) **LT.LIT174:** This subtype A sequence is from a study of HIV-1 in Lithuania. Subtypes A, B and AB-recombinants were found, [Liitsola (1998)]. Accession number AF082481. A gag sequence from this same isolate has accession number AF082448.
- 49) **ML.3665** This subtype A sequence is from a study that looks at the prevalence of different subtypes of HIV-1 and HIV-2 circulating in female sex workers in Bamako, the capital city of Mali [Peeters (1998)]. A total of 176 CSWs were tested and 81 were HIV infected. Of the 81, 63 were infected with HIV-1, 7 were infected with HIV-2 and 11 were dually infected with HIV-1 and HIV-2. HMA assays indicated that 80 percent of HIV-1 infections were with subtype A virus. Only the 9 viruses with ambiguous HMA results were sequenced. Out of these 9 sequences one was subtype A, one was subtype D and 7 were subtype G. Accession number Y14364.
- 50) **MQ1.ID#:** These 2 sequences are from Martinique [Ouka (1998)]. Accession numbers AF062009, AF062010.
- 51) **NG2.NG1935:** This sequence is from Nigeria [McCutchan (1999)]. Subtype G and recombinants were also found. Accession number AF069939.
- 52) **NL1.ID#:** These 13 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. Accession numbers L76877, GH902304; L76903, GH9401488; L76883, GH915200; L76913, GH9501283; L76893, GH929927; L76896, ZR9303337; L76870, RW890388; L76875, RW893568; L76905, TZ9401664; L76881, KE913514; L76891, UG928308; L76887, AO924646; L76889, TZ925825
- 53) **NL2.ID#:** These 6 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions.
- 54) **RE1.ID#:** These 4 sequences are from Reunion Island [Bregel-Pesce (1999)]. Subtype B was also found. Accession numbers Y18590, Y18592–Y18594.
- 55) **RU1.ID#:** This sequence is from Russia (Bobkov et al, unpublished 1996). Accession number U33097.
- 56) **RU2.ID#:** These two sequences are from patients living in Kaliningrad, Russia, [Leitner (1996b)]. RU21 is a 40 year old male heterosexual, believed to have become infected in the Ivory Coast in 1988. RU24 is a 44 year old heterosexual male, believed to have become infected in Democratic Republic of Congo (formerly Zaire). Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. Although RU24 is listed in the publication as unclassified subtype, this sequence clusters in phylogenetic analyses with SAS (U43171) and BJ193 (U61859) which are classified as subtype A. Because of the A-outlier nature of the RU24 env sequence, the authors also sequenced the gag P17 region of this sample, and found that it fell with subtype A (sequence not submitted to the databases). RU21 is a more typical subtype A. Accession numbers U69656 and U69658.

- 57) **RU3.ID#:** These 9 highly related sequences are from a set of 41 sequences from 9 IV drug users living in the Krasnodar Kray, Perm', and Saratov regions of Russia. Uncultured PBMC DNA was PCR amplified and cloned and individual clones sequenced [Bobkov (1997a)]. Although the sequences are similar enough to each other to suggest direct epidemiological linkage, no such linkage, other than the fact that they are all IV drug users, is indicated. Sequences from IV drug users living in Ukraine were also nearly identical (see the UA entry). These blood samples were all collected in 1996. More recent studies of HIV infection among IV drug users in the former Soviet Union have shown that it is very likely that the heroin itself is infected with HIV-1 (subtype A in some cases, subtype A/B recombinant in Kaliningrad) during the processing of the drug [Bobkov (1997a)] and [Liitsola (1998)]. Gag p17-24 and Env V4-V5 were also sequenced. Accession numbers U93611, U93612, U93614, U93616, U93618, U93620, U93622, U93624, U93627, U93629, U93631, U93633, U93635, U93636, U93637, U93638, U93642, U93643, U93645, U93646–U93648, U93650, U93651, U93655–U93661. See also the A/B recombinant HIV-1 recently identified in Kaliningrad [Liitsola (1998)].
- 58) **RU4.ID#:** These three sequences are from Russia and are subtype A in both env and gag [Bobkov (1998a)]. Subtype B and A/B were also found, see the RU entries in the subtype B and unclassified subtype sections. Accession numbers for env are AF051493, AF051494, AF051498, AF051499, AF051500, AF051502, AF051503 and AF051504, and for gag are AF051503, AF051505, AF051506, AF051507. Three other subtype A and one subtype B sequences from the Ukraine were also part of this publication, as was another subtype A env (RU1282) for which a database entry was not created.
- 59) **RU5.ID#:** These 20 sequences are from Svetlogorsk, Belarus, and are subtype A [Lukashov (1998a)]. These sequences, along with the set described by [Bobkov (1997a)], clearly illustrate that the processing of heroin with HIV-infected human blood has resulted in an estimated 100,000 IV drug users becoming infected with a monophyletic strain of subtype A HIV-1. This situation is also observed in Kaliningrad, where a subtype A/B recombinant of HIV-1 was introduced into the drug supply. Accession numbers AF061584–AF061603.
- 60) **RW.SF1703:** This sequence is from Rwandan isolate sf170, a biologically active clone reported to be macrophage-tropic. [Cheng-Mayer (1988)]. See also U61870, AF101116 and AF101124, which are not reported to be from SF170, but are greater than 98% identical to it. From this same isolate; 537 bases of the 5' LTR are in M66534, 619 bases of nef are in M81729 and 508 bases of tat and rev are in M66535. For env, the Accession number is M66533.
- 61) **RW1.ID#:** These are six of seven sequences from asymptomatic individuals from Rwanda sampled in 1992. The seventh sequence (92RW009) has now been sequenced over the complete genome and found to be A/C recombinant. Thus it has been moved to the U_RW section. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf (1994)]; [Osmanov (1994)]; [for HIV Isolation (1994)]. Accession numbers U08630–U08640, U08645, U08647–U08665, U08763–U08766, and U08793–U08794. One of these sequences may be mislabelled in the database; U08634 is labelled as patient 016, but is identical to sequences from patient 008. Sequences from samples 92RW020 and 92RW021 were greater than 98% identical to each other, so only 92RW021 is presented here. 92RW20 and 92RW021 have been reported to be NSI and use the CCR5 coreceptor [Dittmar (1997)]. Other sequences from 92RW021 with accessions U08641–U08644 and U08646 as well as 92RW023 U08719, are all significantly similar to the HXB2R lab strain of HIV-1 and are suspected of being contaminants. The entry with accession number Y14418 is reported to be from Cuba, but is more than 99% identical to 92RW020.
- 62) **RW2.ID#:** These 6 sequences are from Rwanda (Saah, A. unpublished 1994). Many clones from each isolate were sequenced, only one of each is presented here. Accession numbers U23216–U23373.
- 63) **RW3.ID#:** Bex,F. et al. Unpublished. These five sequences are from Rwandan patients with AIDS. The sequencing was done by the EEC Centralized Facility for HIV genome characterization, Georg-Speyer-Haus, Frankfurt, Germany. The complete envelope gene for PVP1 is available from a clone obtained after short-term co-culture on donor PBMCs. Two other shorter sequences of PVP1 env direct from patient PBMCs are also available. Accession numbers L07082–L07091. L07088 and L07089 were withdrawn from the databases by the authors, who felt they may represent PCR artifacts.

- 64) **RW4.ID#:** These 10 sequences, are from 7 infants and 3 mothers in a study of mother-infant transmission in Rwanda [Kampinga (1997), Mulder-Kampinga (1993), Lepage (1996), Mulder-Kampinga (1995), Simonon (1994)]. Mother 566 was apparently infected with both subtype A and subtype C HIV-1, but the env regions from the child clones were all subtype A. See also C_RW1.ID#. Accession numbers for child 566 are Z76160–Z76161, Z76167–Z76168, Z76169–Z76176, Z76233–Z76248, Z76262–Z76273, Z76708–Z76716 and Z76717–Z76724; mother 730, Z76353–Z76362; child 564, Z76074–Z76083; mother 226, Z76046; mother 538, Z76423–Z76462, Z76708–Z76716; child 538, Z76134–Z76143, Z76373–Z76382, Z76393–Z76412; child 074, Z75958; child 082, Z75998–Z76009, Z76650; child 081, Z75959–Z75968; child 618, Z76198–Z76207; mother 618, Z76218–Z76231; mother 439, Z76048–Z76057, Z76064–Z76068 and Z76070, Z75979–Z75988.
- 65) **SE.H4:** This sequence is from an infant, born in Sweden to a woman who was believed to have been infected sexually in Uganda [Contag (1997)]. HIV-1 isolation from the mothers was conducted during the three trimesters, a few days after delivery and 6 months after parturition. The children were born between June 1989 and August 1992 and were followed by virus isolation at birth and at 3, 6, 12 and 18 months of age unless circumstances called for an extra sample. Breast-feeding was denied by all of the women described. See also B_SE5.ID#, C_SE2.ID#, D_SE.H3 and AE_SE.H1. Accession numbers U56274–U56283 and U56328.
- 66) **SE1.ID#:** These 24 sequences are from a study of 270 HIV-1 infected individuals living in Sweden [Alaeus (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 bloods samples were available from 33 men and 42 women, from 15 different African countries. The 24 subtype A sequences were from individuals who were thought to have been infected in Tanzania (SE8538 U76147), Uganda (SE6882 U76140, SE7172 U76118, SE7535 U76154, SE8891 U76160, SE8553 U76149, SE8889 U76170, SE7823 U76132, SE9572 U76182, SE7888 U76119, SE7889 U76143, SE7441 U76137, SE9281 U76175, SE7727 U76125, SE8735 U76159 and SE8131 U76121), Somalia (SE7253 U76165), Kenya (SE8566 U76186, SE8132 U76145, SE7531 U76141 and SE8876 U76168), Ivory Coast (SE7812 U76142), Mauritania (SE9562 U76181) and Mozambique (SE8551 U76151). Two individuals (SE7108 U76131 and SE7019 U76138) were epidemiologically linked to others (SE8566 and SE7531 respectively) and are not included in this alignment. A complete genome of SE8131 is available with accession number AF107771, and shows that this sequence is intersubtype A/D recombinant. SE8538 has been removed from this set, because the complete genome is now available with accession number AF069669 and included in the following set.
- 67) **SE2.ID#:** These 3 sequences are from Sweden [Carr (1999)]. Complete genomes were sequenced. Accession numbers AF069669–AF069673.
- 68) **SG1.ID#:** This sequence is from a study done on HIV-1 sequences from Singapore. Other sequences in this set were subtypes B, C and CRF01(AE), [Se-Thoe (1998)]. Accession number AF004240.
- 69) **SN1.ID#:** These 10 sequences are from a study done on individuals infected with non-B clade virus who were randomly obtained from a cohort of registered sex workers in Senegal West Africa [Cao (1997)]. PBMC were separated, cryopreserved and shipped to USA for CTL studies. Of the 14 sequences evaluated 10 were subtype A, three were subtype G and 1 was subtype C. Accession numbers for the set are AF020819–AF020832.
- 70) **SN2.ID#:** These 9 subtype A sequences are from Senegal [Kanki (1999)]. Twenty-one other sequences which were subtype A in the V3 region were classified as CRF02(AG) circulating recombinant form, because they clustered with the Djibouti and Nigerian AG recombinants within subtype A. Subtypes C, D and G and another AG recombinant were also found in Senegal. Accession numbers for the subtype A sequences are AF085286, AF085287, AF085293, AF085294, AF085299, AF085309, AF085310, AF085319 and AF085326.
- 71) **TW1.ID#:** These 2 sequences are from Taiwan. Another sequence in the set was subtype A/G recombinant, clustering with subtype G in the Gag p24 region [Lee (1998)]. Patient A1 was a 38 year old male sailor who was found to be seropositive in 1989, and was believed to be infected via heterosexual contact. Patient A2 was a 59 year old woman identified as seropositive in 1995 and also infected heterosexually. Accession numbers AF020600 and AF020601. Gag p24 sequences are found with accession numbers AF020948, AF020950.

- 72) **TZ1.ID#:** These 4 sequences are from a set of 15 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing [Zwart (1993)]. The other 11 samples were subtypes C (1) and D (10). Accession numbers L01312–L01316, L01335–L01339.
- 73) **TZ3.ID#:** These 4 sequences are from the Mara region of rural northwest Tanzania [Robbins (1996)]. Subtype D was also found in this study. Accession numbers U61875–U61878.
- 74) **TZ4.ID#:** These 25 sequences are part of a set of 86 sequences from samples collected from symptomatic AIDS patients in December 1995 at Mbeya Referral Hospital in southwest Tanzania. Uncultured PBMC DNA was PCR amplified and directly sequenced. Serotyping was also done on all samples to test the ability of serology to subtype these A, C, D and recombinant HIV-1 isolates [Hoelscher (1997), Hoelscher (1998)]. The sequences have not yet entered the databases (12-23-99).
- 75) **TZ5.ID#:** These sequences are from Dar es Salaam on the eastern coast of Tanzania [Renjifo (1999)]. Subtypes C and D and many independently-derived recombinants were also found in this study. Accession numbers for the entire set are AF038051–AF038121 and AF106332–AF106472. Accession numbers for sequences which are either subtype A in the V3 region and not sequenced in gag, or subtype A in both gag and env are AF038051–AF038060, AF038062–AF038066, AF106332–AF106341 and AF106343–AF106349.
- 76) **UA1.ID#:** These 17 sequences are from a study of molecular epidemiology of an HIV-1 subtype A sub-cluster among injection drug users in Southern Ukraine. Recently another IVDU epidemic in Kaliningrad was found to be caused by HIV-1 that is subtype A in gag and subtype B in env [Liitsola (1998)]. The subtype A gag region was closely related to the subtype A gag of HIV-1 from this Ukraine epidemic. Accession numbers AF025580–AF025596.
- 77) **UA2.ID#:** These three sequences are from Ukraine and are subtype A in both env and gag [Bobkov (1998a)], or were not sequenced in gag. Subtype B was also found in the Ukraine, see the UA entry in the subtype B section. Accession numbers for env are AF051510, AF051511, AF051517, AF051518, AF051421, AF051522 and for gag is AF051523.
- 78) **UA3.ID#:** These 3 highly related sequences are from a set of 41 sequences from 3 IV drug users living in the Donetsk (UA1127 and UA1132) and Mykolaiv (UA1165) regions of Ukraine. Uncultured PBMC DNA was PCR amplified and cloned and individual clones sequenced [Bobkov (1997a)]. Although the sequences are similar enough to each other to suggest direct epidemiological linkage, no such linkage, other than the fact that they are all IV drug users, is indicated. Sequences from IV drug users living in Russia were also nearly identical (see the RU entry). These blood samples were all collected in 1996. More recent studies of HIV infection among IV drug users in the former Soviet Union have shown that it is very likely that the heroin itself is infected with HIV-1 (subtype A in some cases, subtype A/B recombinant in others) during the processing of the drug [Bobkov (1998a), Liitsola (1998)]. Gag p17-24 and Env V4-V5 were also sequenced. Accession numbers U93665, U93676, U93679. See also the A/B recombinant HIV-1 recently identified in Kaliningrad [Liitsola (1998)].
- 79) **UA4.ID#:** These 4 subtype A sequences are from IV drug users in Odessa and Kiev, Ukraine (Nabatov et al. unpublished 1998). Accession numbers AF098953, AF100934, AF100935, AF100939.
- 80) **UA5.UA8:** This subtype A sequence is from Ukraine by Grebenjuk et al 1998. Accession number Y16080. Subtypes B and A/B recombinant were also found.
- 81) **UA6.ID#:** These 3 subtype A sequences are subtype A in env and subtype A in gag [Liitsola (1998)]. The env sequences for Ukrainian isolates are found with accession numbers AF082483, AF082484 and AF082486. Gag genes from the Ukrainian A subtype can be found with accession numbers AF082444 and AF082445.
- 82) **UG.1033:** This sequence is one of several sequences of blood and CSF samples taken from a Ugandan patient 1033, CDC class IV-A. [Keys (1993)]. Accession numbers Z23177, Z23182–Z23184, and Z23220–Z23223.
- 83) **UG.13655:** This sequence is from Uganda (Brooks et al, unpublished 1992). Accession number U04537.
- 84) **UG.92UG037:** This sequence is from a complete genome PCR amplified from proviral DNA. The patient was a 31 year old asymptomatic female from Entebe, Uganda [Gao (1996b)]. Accession number U51190. See also U09124, U09127, U15119, U08769, U08770, U08788–U08792.
- 85) **UG.964:** A single sequence used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from

- peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda [Pestano (1995)]. Accession number U11599. See also B_US17.ID#, C_UG1.45, and D_UG7.ID#.
- 86) **UG.U455:** This sequence is from the 1985 Ugandan isolate U455; the complete genomic sequence is available [Oram (1990)]. The env ORF in this sequence is interrupted by an in-frame stop codon beyond the COOH end of the V5 region. Accession number M62320.
 - 87) **UG.UG018:** A single sequence from Uganda [Buonaguro (1998)]. Accession numbers AF062521 and U44886.
 - 88) **UG.UG06:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was pCR amplified and cloned. This sequence is from an individual clone. [Atkin (1993), Pestano (1993)]. Accession number M98503.
 - 89) **UG1.ID#:** These 2 sequences are from asymptomatic individuals from Uganda in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf (1994)]; [Osmanov (1994)]; [for HIV Isolation (1994)]. Accession numbers L34667, U08666–U08669, U08767–U08770, U08788–U08792, U08795 and U09125. The 92UG037 sequence with accession numbers U08769–U08770 and U08788–U08792 has been previously described with accession numbers U51190, U09124, U09127 and U15119.
 - 90) **UG2.ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic, sampled in 1990. PCR-clones, peripheral blood DNA [Albert (1992)]. Accession numbers M98900, M98902–M98905, M98908–M98910, M98914–M98917, M98919, M98924–M98928, M98938–M98941, M98946–M98966, and M98976–M98978.
 - 91) **UG4.ID#:** Two Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession numbers L22957 and L22951.
 - 92) **UG5.ID#:** These 8 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending a clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4% of the 217 tested) were all asymptomatic [Buonaguro (1995)]. Accession numbers U44878–U44880, U44882, U44883, U44885 and U44887. 94UG018 (U44886) has been previously described with accession number AF062521. Two subtype D sequences were also found in this study (see D_UG8).
 - 93) **UG6.ID#:** These 21 sequences are from Uganda [Ellenberger (1999)]. Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000496–AF000516
 - 94) **ZA.ZA134:** This sequence is from a study of 72 seropositive women from South Africa [Moodley (1998)]. The mean age was 26 years. Patient 134 was asymptomatic. Data from this study shows the dramatic growth of HIV-1 subtype-C in this population in South Africa. See also C_ZA3.ID# and B_ZA.0117. Accession number AF053286.
 - 95) **ZA1.ID#:** These 2 subtype A sequences are from South African sequences published in [Van Harmelen (1999)]. Accession numbers AF095829, AF095830. Subtypes B, C and D were also found in this study in South Africa.

B Subtype

At this time we have included viral sequences from 1910 HIV-1 infected individuals associated with HIV-1 subtype B. The B subtype consensus sequence (B_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. Please note that none of the studies which have published sequences of only the V3 loop sequences are included here, as the DNA sequences were deemed too short for phylogenetic analyses. (For example, LaRosa G, et al., *Science* **249**:932–935 (1990) and Fouchier RAM, et al., *J. Virol.* **66**:3183–3187 (1992).)

- 1) **AR.21281**: This sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient had AIDS and reported promiscuous heterosexual risk behavior. Two other samples taken from unrelated patients in 1993 were subtypes F and one was found to be a subtype B/F recombinant [Marquina (1996)]. Accession number U68525.
- 2) **AR1.ID#**: These 11 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were tested by HMA [Campodonico (1996)]. Accession numbers U37030, U37031 and U37034–U37042.
- 3) **AR2.ID#**: These 12 sequences are from Argentina. A total of 25 patients were included in this study. Of the 25 sequences determined, 12 were found to belong to subtype B and 12 were found to belong to subtype F1, and one was B/F recombinant [Fernandez-Medina (1999)]. Accession numbers AF155514, AF155515, AF155516, AF155518, AF155519, AF155520, AF155524, AF155525, AF155526, AF155531, AF155534 and AF155536.
- 4) **AU.979H** This sequence is a subtype B env sequence from Australia [Wang (1998)]. This paper discusses a patient infected with a rare HIV-1 strain with a cysteine at position 13 within the V3 loop and a GHI insertion on the NH2 side of the loop tip. Accession numbers AF052979–AF053001.
- 5) **AU1.ID#**: Put forth as evidence that coinfection by multiple HIV-1 strains can occur in vivo, these three sequences (UD12, UMR25, and UMR8) come from an Australian homosexual male who had been infected by more than one sexual partner, and harbored three distinct strains of HIV-1 subtype B. The authors also found recombinant sequences, not included here. The sequences were PCR amplified from plasma RNA and PBMC DNA. [Zhu (1995)]. Accession numbers U16372–U16388.
- 6) **AU2.C18CG**: This sequence is one of a group relating to an Australian blood donor infected with HIV-1 and six Australian recipients, all of whom remain symptom free with normal CD4 counts 10 to 14 years after infection [Deacon (1995)]. Samples from only the donor, D36 (U37271), and two patients, C18 (U37267, U37270) and C98 (U37268, U37269), appear to have been sequenced. These sequences have deletions in the nef gene and in the region of overlap of nef and the U3 region of LTR. The authors point to the importance of NEF or the U3 region of LTR in determining the pathogenicity of HIV-1 and suggest this strain of HIV-1 as a possible basis for a live attenuated vaccine. The complete genome of the virus from recipient C18 is in the entry with accession number U37270. Other complete genomes from this set are found with accession numbers AF042100–AF042106.
- 7) **AU3.ID#**: In this analysis of transmitting and non-transmitting mothers and the infants to whom the virus was transmitted, the authors claim that differences in variability of the env V3 loop crown octapeptide (HIGPGRAF in this clone) were significantly correlated with vertical transmission. Sequences were determined by PCR directly from uncultured PBMCs and cloned PCR products. Only seven of the sequences are presented here: from infants, and nontransmitting mothers from which at least 250 bp was sequenced. 1089, 1063, 961 and JW are the infants. Accession numbers for transmitting mothers and their infants are: U66627–U66633, U66638–U66645 and U66650–U66653. Accession numbers for nontransmitting or indeterminate mothers are: U66625, U66626, U66634–U66637, U66646–U66649 and U66654–U66660. Database entries for many of these sequences contain “nnn” in error, the paper shows these positions as deletions, not unknown sequence. Nef sequences from the LW-JW mother-infant pair of long-term survivors were published in [Wang (1997b), Wang (1997a), Saksena (1997)] with accession numbers U73339–U73369.
- 8) **AU4.ID#**: These 4 sequences are from Australia [Naif (1999)]. This study examined viral infection and replication kinetics in macrophages from different human hosts. Monocyte-derived macrophages were harvested from identical and non-identical HIV negative twins, as well as matched non-related control

blood donors. The infection and replication kinetics of different primary virus isolates, as well as the cell line adapted BAL molecular clone, were then tested on the different human cell lines. Sequences of viruses after passage on each of the cell lines were determined, but only one sequence from each of the four primary isolates is presented here. All four isolates came from AIDS patients classified as CDC stage IV with CD4+ T cell counts of less than 50. Accession numbers for the set are AF133342–AF133381. Vpr genes were also sequenced with accession numbers AF133382–AF133390.

- 9) **BB1.ID#:** These 3 sequences are from a study that was done on Barbados patients who tested negative for *Leptospira* infection, implicating other diseases [Roth (1997)]. After doing a survey, 10 HIV-1 positive patients originally hospitalized during 1990-1994 whose medical histories suggested HIV-1 illness at the time of *Leptospira* testing, were found to be HIV-1 infected with symptoms suggesting acute primary infection. Three of the 10 samples were successfully RT-PCR amplified from stored serum RNA. The LL6 PCR product was directly sequenced, the other two were cloned and sequenced. One sequence from each patient is included here. Accession numbers U80239, U80246 and U80247
- 10) **BE.SIMI84** One of two cloned env sequences from a patient with AIDS from Belgium. A vaccinia construct that expresses this gene was created to vaccinate the patient's uninfected brother with the goal of immune therapy by adoptive transfer of lymphocytes [Bex (1994)]. Accession number L07421.
- 11) **BE1.ID#:** These 38 sequences are from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. One of these patients (822) was triple infected with subtypes A, B and E. Accession numbers AJ228190, AJ228166, AJ228189, AJ228155, AJ228151, AJ228161, AJ228141, AJ228170, AJ228184, AJ228145, AJ228162, AJ228164, AJ228187, AJ228171, AJ228178, AJ228202, AJ228188, AJ228181, AJ228149, AJ228153, AJ228177, AJ228172, AJ228147, AJ228174, AJ228168, AJ228160, AJ228146, AJ228173, AJ228223, AJ228192, AJ228154, AJ228179, AJ228144, AJ228140, AJ228143, AJ228191, AJ228158 and AJ228196.
- 12) **BO1.ID#:** These seven sequences are from Bolivia (Velarde-Dunois et al, unpublished 1998). Subtype F was also detected in this study. Accession numbers AF031945–AF031949, AF031951 and AF032901.
- 13) **BR.002:** This sequence is from entries with accession numbers L35489–L35493. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. Ranjbar, S. et al. Unpublished (1994).
- 14) **BR.BHGM19:** This sequence was generated from DNA derived from PBMC HIV-1 ELISA positive male of unknown health status. This sample was classified as subtype F by HMA method, but is subtype B by phylogenetic analysis. Barbosa, E.F. et al. Unpublished (1996). Accession number U46210.
- 15) **BR.BZ** This sequence is from an individual in a Brazilian HIV cohort study. PBMC DNA was PCR amplified in two sequential rounds, and six cloned PCR products were sequenced on both strands. A single clone containing an uninterrupted envelope open reading frame was reported. [da Costa (1995)]. Accession number U28336.
- 16) **BR1.ID#:** These 21 sequences represent the subtype B env sequences found among 22 Brazilian outpatients with varying degrees of disease progression. They are from cloned PCR products. PCR was performed on PBMC DNA [Potts (1993b)]. Accession numbers for 21 of the 56 clones are: L19225–L19236, L19240–L19246 and L20963. One other sequence was subtype F (L19237).
- 17) **BR2.ID#:** These 13 sequences are from individuals from Brazil. They are from cloned PCR products. Some of the clones were from cell-culture DNA, and some from cell-culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database [De Wolf (1994), Osmanov (1994), for HIV Isolation (1994)]. Accession numbers U08670–U08714, U08771–U08778, U08780–U08782, U08792, U08796–U08800.
- 18) **BR3.ID#:** These 32 sequences came from the Brazilian cities Rio de Janeiro and Sao Paulo. The sequences that are very short, containing V3 loop fragments insufficient for phylogenetic analysis, are not included here (5 of the 26). The full set included 20 viral sequences of the B subtype, an F subtype and a B-F recombinant (see subtypes F and U). The year of isolation for the sequences ranged from 1990–1992 for Rio de Janeiro, and 1992 for Sao Paulo. The only two with CD4+ cells < 200 were RJ636 and RJ27. The CDC clinical class ranged from II-IV. DNA extracted from PBMCs of HIV infected individuals was amplified, and the PCR product was directly sequenced [Morgado (1994), Sabino (1994c), Sabino (1996)]. More recent unpublished sequences from the same isolates are also included here. Accession

- numbers U00400–U00401, U00403, U00405, U00407–U00418, U00421, U00424–U00425, U00427, U08975, U31586, U31587 and U31589–U31591.
- 19) **BR4.ID#:** These 2 sequences are from seropositive Brazilian patients. Virus was cultured on donor PBMCs and proviral DNA was harvested from positive cultures. PCR was used to generate sequencing templates. [Louwagie (1994)]. Accession numbers L22087 and L22088. The gag gene sequences from these same isolates are also available in L11752 and L11754. See also F_BR2.BZ-ID#.
 - 20) **BR5.ID#:** These 10 sequences are from entries with accession numbers L19328–L19337 (Banda, C. I. unpublished 1993).
 - 21) **BR6.ID#:** These six sequences are from patients living in cities in Brazil (P3, Sao Paulo; P4, P6 and P7, Bahia; P8, Parana; P9, Rio de Janeiro) and sampled between 1987 and 1989. Sequences were determined from directly sequenced PCR products (except P6 which was a cloned PCR product), after coculture of patient PBMCs with donor PBMCs [Couto-Fernandez (1994)]. Accession numbers X78512–X78517.
 - 22) **BR7.19b:** This subtype B sequence is from Brazil [Janini (1998)]. This study describes a case of horizontal (heterosexual) and subsequent vertical (mother to infant) transmission of 2 HIV-1 subtypes, B and C [Janini (1998)]. DNA sequence analysis of pol, gag and env genes confirmed the presence of subtypes B and C in 3 family members. Accession numbers for env, gag and pol genes of both subtypes are U83689–U83699. Subtype B env accessions are U83689, U83691 and U83693.
 - 23) **BR8.ID#:** These 10 subtype B sequences are from a study of the prevalence of GWG in the V3 loop tip in Brazil [Covas (1998)]. PCR products were first screened for the Proline to Tryptophan variation via RFLP prior to sequencing. Overall prevalence of the GWG sequence as indicated by RFLP was 57% (43 of 75). The prevalence in females (72%) was higher than that in males (32%) and newborns (40%). Accession numbers U80824–U80833.
 - 24) **BR9.ID#:** These 7 sequences are from Brazil (Couto-Fernandez, J.C. unpublished 1999). Subtype F1 was also found in this study. Accession numbers Y18752–Y18755, Y18757, Y18759 and Y18760.
 - 25) **BR10.ID#:** These 4 sequences are from Brazil (Bongertz, V. et al, unpublished 1999). Accession numbers AF060953–AF060956
 - 26) **BR11.ID#:** These 32 sequences are from study of asymptomatic blood donors in Rio de Janeiro Brazil [Tanuri (1999)]. One of the 42 sequences reported was subtype B in env and F1 in gag. Two were subtype F1 in env and B in gag. Six were subtype F1 in both env and gag. One was subtype D in env V3 and gag, but subtype B in the env gp41 region. Accession numbers for subtype B are AF033993–AF034002, AF034004, AF034005, AF034010, AF034012, AF034013, AF034015, AF034016, AF034018, AF034019, AF034021–AF034030 and AF034032–AF034034
 - 27) **BR12.ID#:** These 5 sequences are from Brazil, Vincente, A.C.P. et al. Unpublished (1999). Accession numbers AF076313, AF076316, AF076318, AF076319, AF076322.
 - 28) **BR13.ID#:** These 8 sequences are from Brazil, [Ramos (1999)]. Accession numbers AF113565, AF113569–AF113574 and AF113576.
 - 29) **CA.CAN0:** This sequence is from the Canadian airline steward who was once thought to represent “patient zero”, the person who brought HIV-1 subtype to the USA homosexual community. He had many sexual contacts both before and after developing Kaposi’s sarcoma. Although he may have been the source of infection or “patient zero” for one subcluster of infections, he is not likely to have been the first or only person to import HIV-1 subtype B into north America. This sequence is from serum collected from this patient by the CDC in 1984. Accession number U43103.
 - 30) **CA1.ID#:** These 19 sequences are from 19 different Vancouver, Canada homosexuals represented in 36 database entries by N. Michael et al. Accession numbers U52888–U52906, U53103–U53119. Patients A and B (U52888–U52904) were described in [Michael (1997)] as a normal-progressor rapid-progressor pair, with patient A being the rapid progressor. Only one patient A sequence (U52898) is presented here, because patient B sequences were closely related. The C patients were local controls, also homosexuals from Vancouver, who are not well described in the publication. Patient A sequences from U52901 and U52902 are greater than 99% identical to JR-CSF over their entire length, and U52904 has some regions of identity to JR-CSF and likely represent PCR-recombination artifacts. Another linked pair in this set included patients C108 and C118, of which only C108 is presented here. Patient A died of AIDS-related *P. carinii* pneumonia on October 6, 1995, almost exactly 10 months after experiencing symptoms of primary HIV infection. Patient A was homozygous for for an HLA-DR1 allele previously found to be

associated with rapid progression. Patient A never seroconverted for HIV, despite high viral load and the ability to seroconvert for hepatitis B after vaccination in 1990.

- 31) **CA3.ID#:** These 84 subtype B sequences are all from Vancouver, British Columbia, Canada. Fifty of the 85 samples were from IV drug users enrolled in the Vancouver Injection Drug User Study. Thirty-seven were from homosexual men, some of which were enrolled in a long-term Vancouver study [Estable (1998)]. The IVDU sequences formed several rather tight phylogenetic clusters, indicating rapid spread of the virus among large numbers of people. It is not clear from the publication why the numbers of sequences and samples (84 vs 85 vs 50 + 37) do not quite add up. Patient 6V530 is one of the missing entries. Accession numbers AF058090–AF058173.
- 32) **CH1.ID#:** These 19 sequences came from 24 individuals living in Geneva, Switzerland who were recently infected at the time of blood drawing. Samples were collected between January 1988 and September 1993. Sequences were determined directly from PCR products of uncultured PBMC DNA or serum cDNA. All subjects were asymptomatic, 19 subjects had p24 antigen levels ranging from 5 to 6,357 pg/ml and 5 subjects had no detectable p24 antigen. Two subjects were epidemiologically linked (K11 and K16) so only one of those two is presented here. Two other individuals showed identical DNA sequences over the entire V3 region (K53 and K77) so only one of them is presented here. Three other individuals (K13, K42 and P4) had sequences nearly identical to the LAI (IIIB) laboratory strain of HIV. Although the authors are convinced that these are not contaminants, and that a IIIB-like strain of HIV is circulating in Geneva, they are not included in this alignment [Antonioli (1995)]. Accession numbers U10957–U10980.
- 33) **CI.CI-22:** A single B subtype sequence from a set of 13 isolates from individuals from Abidjan, Ivory Coast. CI-22 was symptomatic. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3 clones were sequenced, and one of the clones is presented here. [Janssens (1994a)]. Accession numbers X72040–X72042.
- 34) **CM.CA5:** A single B subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, patient CA-5 was asymptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate [Nkengasong (1994)]. Accession number X80452.
- 35) **CM.277** This sequence is from a 1994-1995 study of 211 Cameroonian AIDS patients [Takehisa (1998)]. Of the 43 HIV samples sequenced, 17 were subtype A, 1 was subtype B, 2 were subtype C, 1 was subtype G. Accession number AF023083. A pol gene sequence from this same isolate is found with accession number U69228.
- 36) **CM.909:** This sequence is from Cameroon (Roques et al, unpublished 1998). Accession number AJ006741.
- 37) **CN.1798:** This sequence is from a dried blood spot collected in 1992 from the spouse of an IV drug user in China. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol (1996)]. Accession number U53316.
- 38) **CN.RL42:** This sequence is from a complete genome from an asymptomatic IV drug user in China [Graf (1998)]. Accession number U71182.
- 39) **CN1.ID#:** These 54 sequences are from the Yunnan Province, China (Shao, Y. and Wolf, H. unpublished 1995). Several of these sequences (U20009, U20012, U20013, U20018, U20023, U20024) are greater than 96% identical to the SF2 strain of HIV-1 (see B_US.SF2) and are not included here. Accession numbers U20001–U20054.
- 40) **CN2.ID#:** These 12 sequences are from the Guangxi Province, China from blood samples collected in 1996 [Chen (1999)]. Subtypes C and CRF01(AE) were also identified in this study. Accession numbers AF080186–AF080191, AF080205, AF080206, AF080209–AF080212.
- 41) **CU.94CU053:** This sequence is from a bisexual male, most probably infected via heterosexual contact in 1992, in Cuba. Virus was isolated in 1994, 2 years after seroconversion, by cocultivation of patient PBMCs with donor PBMCs. This isolate exhibits a rapid/high, syncytium-inducing phenotype. [Gomez (1996)]. Accession number U48855.
- 42) **CU1.ID#:** These sequences are from Cuba (Gomez, C.E.G.R. unpublished 1997). Accession numbers Y14411, Y14413, Y14416, Y14417, Y14419, Y14421.

- 43) **CY1.HO#**: These 15 sequences are from samples, like others in this study (see also subtypes A, C, F and CRF04_cpx), which were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR amplified and cloned. Individual clones were sequenced. Patient 04 was a 24 year old man, asymptomatic with a CD4 count of 516, who had lived abroad, and had been seropositive for at least 4 years. Patient 11 was a 29 year old homosexual man, symptomatic with a CD4 count of 277, and seropositive for at least 7 years. Patient 21 was a 38 year old bisexual man, asymptomatic with a CD4 count of 743, who was infected in Greece and seropositive for at least 4 years. Patient 25 was a 34 year old heterosexual man, asymptomatic with a CD4 count of 650, who had lived in the U.S. and was seropositive for at least 6 years. Patient 27 was a 20 year old homosexual man, asymptomatic with a CD4 count of 709, infected in Cyprus and seropositive for at least 2 years. Patient 28 was a 39 year old homosexual woman, asymptomatic with a CD4 count of 430, infected in Cyprus and seropositive for at least 1 year. Patient 29 was a 49 year old heterosexual man, asymptomatic with a CD4 count of 420, infected in Cyprus and seropositive for at least 2 years. Patient 39 was a 37 year old heterosexual man, asymptomatic with a CD4 count of 470, had lived in Greece and was seropositive for at least 5 years. Patient 40 was a 29 year old heterosexual woman, symptomatic with a CD4 count of 92, who had lived in the U.S. and was seropositive for at least 6 years. Patient 43 was a 36 year old homosexual man, symptomatic with a CD4 count of 396, who had lived in the U.K. and Greece and had been seropositive for at least 6 years. Patient 45 was a 32 year old heterosexual man, asymptomatic with a CD4 count of 453, who was infected in Cyprus and seropositive for at least 1 year. Patient 46 was a 32 year old heterosexual man, whose partner had died of AIDS, asymptomatic with a CD4 count of 107, had lived in the U.K. and was seropositive for at least 4 years. Patient 48 was a 22 year old hemophiliac man, asymptomatic with a CD4 count of 276, who had been seropositive for at least 11 years. Patient 50 was 32 year old homosexual man, whose partner had AIDS, asymptomatic with a CD4 count of 315, who had been seropositive for at least 1 year. Accession numbers U28663, U28666 (04); U28664 (11); U28667–U28671 (21, 25, 27, 28, 29); U28675, U28676 (39, 40); U28678 (43); U28680–U28682 (45, 46, 48); U28684 (50).
- 44) **CZ.BTSPR**: This sequence is from a 32 year old homosexual man from Prague, Czech Republic. It was included as part of a study of HIV-1 sequences from the neighboring Slovak Republic. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in the database. It is not stated whether each single sequence is a consensus of the eight clones or a single clone. [Zachar (1996b)]. Accession number U53195.
- 45) **CZ1.ID#**: These 21 sequences are from the Czech Republic [Quinones-Mateu (1999)]. Accession numbers AF080133, AF080134 and AF080136–AF080154. See also CRF01_AE one sequence.
- 46) **DE.D31**: This sequence is from isolate D31. [Kreutz (1992)]. It has never been well described, it is only shown as HIV1-D31 in figure 3 of the paper. The complete genome has been sequenced. Accession number U43096.
- 47) **DE.HAN**: This sequence is from an infectious clone from the German isolate DE.HAN-2 [Sauermaun (1990)]. Isolate HAN2 was isolated from a 39 year old homosexual German patient with AIDS related complex in 1986. This patient died from complications of AIDS in 1987. HAN2 was highly cytopathic in T cell line MT-2 it was able to productively infect MT-4, H9 or Jurkat cell lines. Genomic DNA from infected MT-2 cells was used to prepare a lambda phage genomic library. Two full-length clones, HAN2/2 and HAN2/3 were purified. HAN2/3 was used for DNA sequencing. Accession number U43141.
- 48) **DE.Seroconv**: This sequence is one of several sequences from 7 hemophilia patients who all received the same lot of beta-propiolactone and UV-light inactivated clotting factor in Bonn or Goetingen, Germany, from November 1989 to March 1990. The virus and the patients have been extensively studied over time, since initial seroconversion. The sequences were from proviral DNA from cultured PBMCs, PCR amplified and cloned [Kasper (1994)]. Accession numbers S76444 and S76446.
- 49) **DE1.ID#**: These 7 sequences are from a Neurology thesis by I. Weber of Rostock, Germany. They are from blood and CSF specimens from 6 patients. For each of the patients, phylogenetic analysis showed that blood and CSF sequences were more similar to each other than to other database sequences. For one patient, one of the two CSF sequences (Z48786) was different enough from another blood and CSF pair (Z48782, Z48785) to be included here. Accession numbers Z48782 and Z48785-Z48793.
- 50) **DE2.ID#**: These 4 sequences are from a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and 6

- Scottish intravenous drug users. The sequences were used in a study of HIV-1 vpr, vpu, and env V3 regions and how they vary between risk groups [Kuiken (1996b)]. Accession numbers Z68529, Z68530, Z68537 and Z68538.
- 51) **ES.S61:** This sequence is from a 1989 blood sample from a 4 year-old boy from Madrid, Spain with CDC stage P2CD2 disease. Virus was cocultured on PBMCs and MT-2 cells prior to sequencing. The SI/NSI phenotype of this isolate on MT-2 cells was traced to a single amino acid change in the V3 loop [Sanchez-Palomino (1993)] and [Olivares (1997)]. Accession numbers L04604, L04605, L04606 and L05659. A complete genome has been sequenced with accession number AJ006287 [Neilson (1998)].
 - 52) **ES1.ID#:** These 36 sequences are from 41 patients sampled in Madrid, Spain between 1985 and 1991. Proviral DNA was extracted from uncultured patient PBMCs and the C2V3 region was PCR amplified. The PCR products were directly sequenced. Two of the sequences reported in this set (D22-28 and D22-48) were 99.5% identical to the LAI (IIIB) lab strain of HIV-1 and are not included here. Three other groups of sequences had members that were greater than 98% identical to each other (R1, R2 and R3; THF13-2, THF12-24; S1, S4) and only one of each of them is presented here. [Quinones-Mateu (1996)]. Accession numbers U40533–U40552 and U45286–U45307.
 - 53) **ES2.ID#:** These 21 sequences are from hemophilia patients sampled in Spain [Ortiz (1998)]. Accession numbers AF075722–AF075746.
 - 54) **FR.FR70133:** This sequence is from France (Roques et al, unpublished 1998). Accession number AJ006748.
 - 55) **FR.J91:** This sequence is from one of the JBB clones from the French patient Bru [Wain-Hobson (1991)] and [Guo (1991)]. Accession numbers X57449–X57459 and X57461.
 - 56) **FR.LAI:** This sequence is from the French isolate LAI (formerly BRU) which is also referred to as IIIB [Wain-Hobson (1985)]. Also see: [Alizon (1986), Lukashov & Goudsmit(1995), Wain-Hobson (1991)]. Accession numbers K02013, L23090–L23101, X01762, L48380–L48399, M64178–M64223, M64406–M64415 and M64768–M64775. Other sequences which are of this type include: PV22, K02083; MFA, M33943 [Stevenson (1990)]; F12, Z11530; BH8, K02011; BH10, M15654; TH4, L31963; MCK1, D86068; PM213, D86069; LL13, U80242; H9, L42371; and HXB, AF033819, K03455, M38432 and M14100. The variation of the IIIB isolate in culture was studied by [Lockey (1996)], Accession numbers U54647, U54649, U54651, U54653, U54655, U54657, U54659, U54665, U54667, U54681, U54683, U54685 and U54689. The variation of IIIB/LAI in 9 years of infection in a chimpanzee has been studied by [Wei & Fultz(1998)], Accession numbers U56866–U56883 and U56888–U56899. The IIIB/LAI isolate of HIV-1 has also been extensively studied in cases such as the infected lab worker. See for example [Reitz Jr. (1994), Pincus (1994)] U12030–U12055. The tropism of isolates from the lab worker for primary PBMCs and failure to grow in T-cell lines was localized to the V3-loop by Lishan Su et al. [Su (1997)]. Recombinant virus pNL4-3, with envelope from LAI(BRU) and gag-pol from NY5 has also been studied: [Adachi (1986)] Accession numbers M19921, AF070521, AF003888 [Duensing (1995)], accession number L42371 and [Salminen (1995)] accession number U26942. Other database entries with IIIB-LAI sequences can be found in the patented sequences section of GenBank, in the cloning vector section, and in the primate section (for example U19867, A00647, A04321 and M18404).
 - 57) **FR.PA14:** This subtype B sequence is from a case in France, where the parents of a 31 year old AIDS patient became HIV-infected despite having only casual contact with the patient. There were no needle sticks or other obvious routes of blood contact [Belec (1998)]. Accession numbers for the set are U87171–U87221, and AF092953–AF092961 including V3 and V2-V4 region sequences.
 - 58) **FR1.ID#:** These 4 sequences are from Toulouse, France. In this study, 4 mother-infant pairs were followed during pregnancy and after birth. The inter- and intra-patient sequence similarities of this set of 308 sequences has been controversial, because some infant sequences were identical to sequences from other mothers. For purposes of this V3 section, only one sequence from each of the 4 infants is presented here [Briant (1995), Korber (1995), Jr (1996)]. Database entries for all 308 sequences are found with accession numbers U24717-U24999 and U25001-U25025.
 - 59) **FR2.ID#:** These 20 sequences are from members of the French military who are for the most part believed to have been infected while deployed outside of France (Chad, Djibouti, Gabon, Guyana, Mayotte and Cameroon). An additional sequence (FRMP040, U58787) was listed as subtype C in the paper, but clustered with subtype B in analysis done at the HIV Database and it has been listed as subtype unclassified until more information is available. Other sequences from this study were subtypes A, C,

- E, F, and unclassified [Lasky (1997)]. Accession numbers for subtype B were U58808–U58826 and U58777.
- 60) **FR3.ID#:** These 11 sequences are from individuals from France with primary HIV-1 infection during the peak of viremia [Ataman-Onal (99)]. All of these are subtype B sequences. Two of the patients (160 and 384) showed significant phylogenetic linkage, but 160 was infected in 1993 and 384 was infected in 1996 and the authors reported that they were not known to be epidemiologically linked. Thus all 11 sequences are reported here. Accession numbers AF041125–AF041135.
 - 61) **FR4.ID#:** These 2 subtype B sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers Z95455, Z95456 and Z95460. Z95455 and Z95456 were nearly identical to each other, so only Z95455 is presented here.
 - 62) **FR6.ID#:** These 3 subtype B sequences are from mother-infant transmission cases in France [Pasquier (1998)]. Child CF was born by cesarean section and was HIV-infected and died of AIDS before one year of age. The other 2 subtype B infected infants AZ and BO were born via vaginal delivery and survived more than one year. A fourth mother-infant pair was infected with subtype A virus. Many of the sequences were partly mislabelled in the GenBank submissions. For example AJ009101 is described as “isolate MAZ, DNACBO11” and it is clear that the sequence is related to the BO pair and not the AZ pair. MAZ would be the code for mother AZ, and CBO would be the code for child BO. Accession numbers for subtype B samples are AJ008670–AJ008752, AJ008857–AJ008962, AJ008986–AJ009050 and AJ009072–AJ009113.
 - 63) **GA.OYI:** This sequence is from the Gabonese isolate OYI (designated elsewhere as isolate 397), isolated from a healthy HIV-1 infected individual. GA.OYI appears to have been the first viral sequence from Africa that phylogenetically clustered with North American viruses [Huet (1989)]. Accession number M26727.
 - 64) **GB.AIT:** This sequence is from an individual at the time of seroconversion. Proviral DNA was extracted from PBMCs from a patient who was viremic and had an equivocal HIV-1 antibody status, and the env V3 region was PCR amplified. The PCR products were cloned and the DNA sequence determined for 15 clones. These data showed that the V3 region contained only limited sequence heterogeneity with a major variant (shown here) accounting for 66% of the protein quasispecies present [Ait-Khaled & Emery(1993)]. Accession number S69598.
 - 65) **GB.BM1** This sequence is from a study of person to person transmission [Wade (1998)]. Accession numbers AF050769–AF050892. Gag genes from these isolates have also been sequenced, see accession numbers AF014163, AF050911 and AF050913 for examples.
 - 66) **GB.CAM1:** This sequence is from the British isolate CAM1. McIntosh A, and Karpas A, Thesis (1991), Cambridge University, England. Accession numbers D10112, D00917.
 - 67) **GB.GB8** This subtype B sequence is from a study in which entire env gene (gp160) of GB8, a British HIV-1 isolate, was cloned, sequenced and aligned with other reference strains [Vella (1995)]. Accession number Y13716.
 - 68) **GB.Man** This sequence was PCR amplified from the 1959 “Manchester sailor” kidney tissue. The sequence of the complete genome is available and clusters with subtype B contemporary HIV-1 sequences [Zhu & Ho(1995)]. Using mitochondrial DNA analysis, the tissue was determined to not belong to the “Manchester sailor”. Accession number U23487.
 - 69) **GB.MB314:** This sequence is from a British isolate [Douglas (1997)]. This sequence was part of an analysis of the structure/function of HIV1, HIV-2/SIV-SMM and SIV-AGM env proteins. The study includes useful alignments of various subtypes of HIV-1, as well as HIV-1 env aligned to HIV-2 and SIV env proteins, color coded and with glycosylation sites annotated. Accession number Y13719.
 - 70) **GB1.CPHL1:** This is a sequence from the British isolate 93–08020. It was referred to as 93–08020 in [Arnold (1995c)] and was isolated from the patient referred to as CPHL1 in [Arnold (1995a)]. CPHL1 is a surgeon and CPHL2 was a patient of his in 1986, approximately 7 years prior to sampling for this study. Because sequences from CPHL1 and CPHL2 are no more similar to each other than to sequences from the general population, transmission cannot be concluded, and both sequences are included in this alignment. Accession numbers U21100 (clone 1) and U23112–U23116 (clones 18, 19, 4, 43 and 7 respectively).

- 71) **GB2.CPHL2:** This is a sequence from the British isolate 93–17305 [Arnold (1995c)]. It was isolated from the patient referred to as CPHL2 in [Arnold (1995a)]. Accession numbers U23117–U23120 (clones 11, 18, 25 and 3 respectively).
- 72) **GB3.ID#:** The CPHL7 sequence is a sequence from the British isolate 94–24612, clone 13 [Arnold (1995c)]. It was isolated from the patient referred to as CPHL7 in [Arnold (1995a)]. Accession number U23126. U23127 is a second clone from this same isolate. Sequences from three other patients epidemiologically linked to CPHL7 (CPHL6, accession numbers U23130–U23132; CPHL8, U23128–U23129; CPHL9, U23133–U23135) are not included in this alignment. The 4995 sequence is from entries with accession numbers U23136–U23138.
- 73) **GB4.ID#:** These eight sequences are from British isolates from St. Bartholomew's Hospital, London (M23470, M26864, M30156, M37677 and M37658) and Hammersmith Hospital, London (AC, JB and WB). Sequences were determined from cloned PCR products from PCR amplified DNA from either cocultured (M23470 and M26864) or uncultured (M30156, M37677, AC, JB and WB) patient PBMC proviral DNA [Douglas (1996)]. Complete envelope gp160 sequences were determined for at least one clone from each patient. Ugandan samples also sequenced in this report were subtypes D or D/A recombinant. Accession numbers for London samples were U36859–U36864, U36869, U36870, U36872–U36880 and U36882.
- 74) **GB5.ID#:** These 11 sequences are from Scotland. They are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors found consistent differences in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication [Kuiken (1996b)]. See also B_NL12 and B_DE2 sets.
- 75) **GB6.ID#:** These 5 subtype B sequences were obtained from a study done on 15 individuals. Eleven of the specimens were from heterosexuals, two were from injecting drug users and one was from a homosexual. Two specimens were from one woman whose risk behavior was not known, and who seemed to be dually infected with subtype B and the CRF01(AE) circulating recombinant form. The specimens were collected in England from individuals whose history indicated that they had become infected in Southeast Asia, particularly Thailand [Belda (1998)]. Accession numbers for the sequence set including gag regions are AJ224176–AJ224200.
- 76) **GB7.ID#:** These 2 sequences are from soon after seroconversion. The study noted mutations in the env gene that contributed to coreceptor specificity [Lewis (1998)]. Accession numbers AJ007943–AJ007948.
- 77) **GB8.ID#:** A set of six sequences from a study of hemophiliacs from Scotland who were originally thought to have been infected by the same batch of factor VIII. (ScV12 is a sequence from a hemophiliac from the U.S., included as a control). All are consensus sequences of multiple direct PCR sequences obtained from limiting dilution of PBMCs. The Scottish hemophiliacs were infected in 1984 and the PBMCs were obtained for analysis in 1989. Although the samples were potentially related, they were deemed sufficiently divergent in this region for inclusion in this set [Simmonds (1990), Balfe (1990)]. Accession numbers M61327–M61346 and M61391–M61407. Database entries with accession numbers M84240–M84317 are more sequences from patient 82, taken over the period from 1984–1991. [Holmes (1992), Leigh Brown & Cleland(1996)]. Database entries with accession numbers L13488–L13497 are also from these patients [Zhang (1993)], as are U58393–U58465 [Cleland (1996), Ludlam (1985)]. Gag p17 genes from the same population were published in [Leigh Brown (1997)].
- 78) **GM.GM6:** A sequence from Gambia (Bobkov et al, unpublished 1996). Accession number U33101. See also Gambian sequences of subtypes C and J.
- 79) **GR1.ID#:** These 8 sequences are from northern Greece. The sequences contained numerous frameshifting insertions, possibly due to sequencing errors, [Papa (1998)]. Accession numbers AJ224948–AJ224955.
- 80) **GR2.ID#:** These 35 sequences are from Greece [Nasioulas (1998)]. Subtypes A, C and D were also noted in this study. Many of these sequences had frameshifting insertions and/or deletions, possibly due to sequencing errors. Accession numbers AF049290, AF049291, AF049293, AF049302–AF049319, AF049323–AF049326, AF094522–AF094530 and AF053613.
- 81) **HN1.ID#:** These 17 sequences are from Honduras [Blackard (1999)]. All sequences in this study were subtype B. Accession numbers AF096673–AF096689.

- 82) **HT.H121:** This sequence is from Haiti (Kalish et al, unpublished 1998). Accession number U04535.
- 83) **HT1.ID#:** These seven sequences are from Haitians, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratories of Dr. Beatrice Hahn at the University of Alabama, and Dr. Marcia Kalish at the Centers for Disease Control, Atlanta, GA. Except for D2HA590, the full gp160 was sequenced from clones derived from expanded culture stocks. D2HA590 is a direct sequence from PCR amplified DNA from expanded culture. The sequence ID numbers are abbreviated, for example D2HA590 can be read as DAIDS sequence (D), isolated in 1992 (2), Haitian (HA), patient 301590 (590). Full length env for some of these have been expressed [Gao (1996a)]. Accession numbers: U08441–U08447, U04900. Both U08441 and U08442 are sequences from patient HT1.D1HA651 and are identical over the region of interest. Accession numbers for additional clones derived from these patients: U04901–U04906.
- 84) **HT2.ID#:** These 25 sequences are from Haitians. All sequences were PCR amplified from the infected individuals PBMCs, and this set includes direct sequences of PCR amplification products, consensus sequences of multiple clones of PCR products plus one direct sequence, and single clones of PCR products. Full length env for some of these have been expressed [Gao (1996a)]. These sequences were provided by the Centers for Disease Control, Atlanta, GA USA (Dr. Chin-Yih Ou), and John Hopkins University School of Hygiene and Public Health, Baltimore, MD USA (Dr. Neal Halsey), and the Centers for Development and Health, Complexe Medico Sociale de la Cite Soleil, Port-au-Prince, Haiti (Dr. Reginald Boulos). Accession numbers L07145–L07161, L07163–L07165, L07167–L07207, L07209–L07239, L07241–L07246, U08441–U08447.
- 85) **ID.1701:** This sequence is from a dried blood spot collected in 1992 from a male homosexual patient in Indonesia. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol (1996)]. Accession number U53317.
- 86) **ID1.ID#:** These 7 sequences are from a set of 14 sequences from Indonesia [Porter (1997)]. Accession numbers U68195–U68201. Circulating recombinant form CRF01(AE) was also identified in 7 other samples from Indonesia in this study. PCR products were directly sequenced from either uncultured PBMC DNA or cocultured PBMC DNA.
- 87) **IL1.ID#:** These 7 subtype B sequences are from Israel [Gehring (1997)]. Two other subtype B sequences from this study were too short to include here. Other subtypes found in Israel in this study were A, C, D, K and CRF02(AG). Accession numbers for the subtype B sequences are X94376–X94378, X94381 and X94383–X94385.
- 88) **IN.IND9:** This sequence was from a heterosexually infected patient from New Delhi, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced [Tripathy (1996)]. Accession number U31364. See also C_IN3.ID#.
- 89) **IN1.ID#:** These four sequences were isolated in Hyderabad, Andhra Pradesh, in southern India. The C2V3 region of env was amplified by nested priming from DNA from PBLs from fresh blood samples. Date of sampling and health status of HIV-1 infected individuals is unknown [Baskar (1994)]. Accession numbers L29091–L29094.
- 90) **IN2.ID#:** These 2 sequences are from India. Another sequence from this publication was subtype A (see A_IN3) and seven were subtype C (see C_IN6) [Maitra (1999)]. The subtype B sequences were from IV drug users from Manipur in northeastern India, and they were of the B-prime sub-clade of subtype B. Another IV drug user from Manipur had one of the subtype C sequences. Accession numbers AF101121 and AF101122.
- 91) **IN3.ID#:** This sequence is from India. Another sequence (26840, AF148227) is subtype B/C recombinant. A third subtype B sequence in this set (47398, AF148262) was not included here because it is very similar to 47399 (Seth et al, unpublished 1999). Other sequences in this set were subtype A and C (see A_IN4 and C_IN7). Accession numbers AF148262, AF148263 and AF148227.
- 92) **IT.PD:** This sequence is from the thymus of a 26 year old male with CDC stage III disease. Proviral DNA was PCR amplified from genomic DNA extracted from patient thymocytes, and the PCR product was directly sequenced. A V3 sequence from PBMC proviral DNA was also determined, as were gp41 sequences from both thymus and blood [Calabro (1995)]. Accession numbers U09254–U09255.
- 93) **IT.RM1:** This sequence is from a patient from Milan, Italy who was homozygous for the delta-32 deletion in CCR5 [Balotta (1997b)]. Accession number U92491.

- 94) **IT1.ID#:** These 10 sequences are from infants infected in utero. The sequences came from PCR amplified DNA of uncultured PBMCs, PCR amplified DNA of cultured PBMCs, or from RNA from serum collected at or shortly after delivery. [Scarlati (1993)] and [Halapi (1997)]. Accession numbers L08277–L08286. Sequences from the mothers of these infants are also available in entries with accession numbers L08287–L08372 and more sequences are available with accession numbers AF023344–AF023419.
- 95) **IT2.ID#:** These 3 sequences are from sequences used to express HIV-1 env gp120 on the surface of Streptococci [Oggioni (99)]. DNA was amplified from uncultured PBMC and cloned prior to sequencing. Health of the patients was not specified. Accession numbers X92424–X92426.
- 96) **IT3.ID#:** These 12 sequences are from 12 individual patients in a study of long term non-progressors and rapid progressors, from Milan, Italy [Balotta (1997a)]. Accession numbers U95381–U95497.
- 97) **IT4.ID#:** These 10 subtype B sequences are from a set of 450 subtype B sequences from a mother-infant study in which maternal samples were collected within seven days of delivery [Salvatori (1997)]. Peripheral blood samples from the children were supplied by the Pediatric department of University of Padova. All of children were full term infants born by spontaneous vaginal delivery; none were breast fed. The children were followed clinically and immunologically every month during the first three months of life and then every 2-3 months. One sample from each infant is included here. Accession numbers U75185, U75101, U75081, U75016, U74997, U74922, U74900, U74841, U74819 and U74772
- 98) **IT5.ID#:** These 6 subtype B sequences are from a set of 278 subtype B sequences from a study of inpatient variability over time [Bagnarelli (1999)]. All six patients were IV drug users who were asymptomatic (four males and 2 females). Patients A, B and C were typical progressors. Patient D, E and F were slow progressors. Patient E was a sex partner of patient B, but the sequences from the two patients do not seem to be related, most likely because both were infected via IV drug use, rather than sexually. Accession numbers AF105433–AF105594 and AF105846–AF105961. Some of the entries from patient D are more than 94% identical to the BAL isolate, and might possibly represent PCR contamination events, but it has also been noted that the BAL isolate from USA has all the “signatures” of a European IVDU sequence. PCR contamination most often results in sequences with greater than 98% identity. The BAL isolate also shows similar levels of identity to other Italian IVDU sequences.
- 99) **IT6.ID#:** These two sequences are from 2 patients with 4 clones each. They were obtained from PCR-amplified proviral DNA from Langerhans cells from skin patches from a deceased AIDS patient in Italy [Sala (1995)]. Small V1-V2 region sequences and V3-loop sequences from the same skin samples were published in [Sala (1994)]. Entries with accession numbers U20670–U20677 were used here. Entries with accession numbers Z34376–Z34458, Z34470–Z34513 and Z34515 were V1-V2 and V3-loop sequences from the same patient.
- 100) **JP.ETR:** A Japanese isolate from long-term cell culture with a truncated env gene, due to a point mutation of a CAG codon to a TAG stop codon [Shimizu (1992)]. Accession numbers D01205–D01207, D12582, D12584 and D12571.
- 101) **JP.GUNA:** A Japanese 1989 isolate HIVGUN, infectious to T cells, was adapted to grow in fibroblast-like BT cells. A single amino acid change at the tip of the V3 loop was shown to be responsible for the change in tropism, GPGR to GSGR. [Takeuchi (1991)]. Accession numbers M59192 and M59193.
- 102) **JP.JH23A:** This sequence is from a Japanese patient dually infected with HIV-1 subtypes B and CRF01(AE) [Xin (1995a)]. Accession number D67090. See also D67089, AE_JP.JH23B.
- 103) **JP.JH32:** This is a sequence from a lambda clone of Japanese isolate JH3, which was isolated in 1986 from a 10 year old Japanese Hemophiliac. [Komiyama (1989)]. Accession number M21138.
- 104) **JP.JNIH1M:** This sequence is from a Myanmarese (Burmese) individual living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells [Weniger (1994)]. Accession number L32084.
- 105) **JP.KM03:** This sequence is from a 28 year old hemophilia B patient with CDC stage IV disease and T-cell count of 20, living in Japan. The authors [Hattori (1991)] also sequenced the V3 region from 28 other Japanese individuals, but only the V3-loop amino acid sequence is available from the other patients. Accession number S70936.
- 106) **JP1.ID#:** These 10 sequences are from a study of 12 patients with varying rates of disease progression [Shioda (1997)]. Patients 1 and 2, both extremely rapid progressors who died of AIDS within 8 months (patient 1) and 3 years (patient 2) of infection by sexual contact in 1991. These two shared very similar sequences and were suspected to be epidemiologically linked, so only patient 2 sequence is presented

- here. Patients 8, 9, 15, 19, 47, 63 and 65 were infected between 1983 and 1985 by contaminated blood products. Patients 20 and 28 were infected by sexual contact. Sequences were determined by either RT-PCR from plasma viral RNA or PCR from uncultured PBMC proviral DNA. PCR products were cloned prior to sequencing. Each of these 11 sequences is a single clone, taken as representative of that patient. Accession numbers AB002829–AB002974, AB002988–AB003019. Patient 43 was infected with CRF01(AE).
- 107) **JP2.ID#:** These 4 sequences are from 5 hemophiliacs from Japan. Two of the 5 (AB010410 and AB010412) were very similar to each other so only AB010410 is included here [Okamoto (1998)]. Accession numbers AB010409–AB010413.
 - 108) **JP3.ID#:** These 20 sequences are from a study of 26 blood donors from Japan [Kitsutani (1998)]. The other six were of the Thai CRF01(AE) form. Although several pairs of the sequences (J7, AB012968 and J12 AB012973; J19 AB012980 and J25 AB012986; J13 AB012974 and J24 AB012985) were quite similar to each other, both of each pair are included here, because there was no evidence of direct epidemiological linkage. Accession numbers for the entire set are AB012962–AB012994. Accession numbers for these 20 subtype B are AB012962–AB012987 and AB012994
 - 109) **KR.Kr111:** This sequence is from an unpublished database entry with accession number X93580.
 - 110) **KR1.ID#:** These sequences are from South Korea. Subtypes A, C and H were also reported, but the subtype H sequence was not submitted to the databases [Kim (1999)]. A subtype A/F recombinant (KR51) was submitted to the database, but not reported in the publication. Accession numbers for the entire set are Z92548–Z92668.
 - 111) **LB1.ID#:** These 10 subtype B sequences are from a study of 25 sequences from a study of multiple HIV-1 subtypes in Lebanon [Pieniazek (1998)]. Of the 26 samples, 25 were classified as HIV-1: 10 were subtype B, 10 were subtype A, 1 was subtype C, 1 was subtype D and 3 were recombinants or untyped. The other sample was classified as HIV-2 subtype B. Accession numbers AF025692, AF025694, AF025696, AF025698, AF025701, AF025702, AF025704, AF025706, AF025708 and AF025714 are HIV-1 subtype B.
 - 112) **LT1.ID#:** These 5 sequences are from Lithuania. [Lukashov (1995)]. Accession numbers: L38417, LIT18A; L38412, LIT11A; L38419, LIT21A; L38416, LIT17A; L38420, LIT17A.
 - 113) **LT1.ID#:** These 2 sequences are from Lithuania [Liitsola (1998)]. Accession numbers: AF082479, AF082482. Subtype A and AB-recombinant were also found.
 - 114) **MM1.ID#:** These two sequences are from Myanmarese (Burmese) individuals living in Myanmar, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells [Weniger (1994)]. Patient 02 is a male IV drug user, and 05 is a female prostitute, both were from Mandalay. Accession numbers L32088, L32089.
 - 115) **MM2.ID#:** These 5 sequences are from dried blood spots collected in 1992 from a male STD patient (1782), a female prostitute (1739), and 3 IV drug users (1748, 1755 and 1763) in Myanmar. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol (1996)]. Accession numbers U53304–U53308.
 - 116) **MM3.ID#:** These 27 sequences are from a set of 30 patient samples sequenced from Yangon, Myanmar. Serological subtyping was performed on blood samples from other regions of Myanmar in the same study [Kusagawa (1999)]. IDU in Yangon were all subtype B' as were some of the heterosexually acquired infections. Circulating recombinant form CRF01(AE) was not found in any IDU in Yangon, but only in heterosexual cases. DNA was extracted from PBMCs and PCR amplified. The PCR products were directly sequenced. All samples were collected in 1995, and all but 6 of the 30 patients had AIDS. Accession numbers AB010747–AB010754, AB010756–AB010764, AB010766–AB010775.
 - 117) **MQ1.ID#:** These 11 sequences are from Martinique [Desgranges (1996)]. Subtype F was also found. Accession numbers for subtype B are U67701–U67706 and U67711–U67715.
 - 118) **MY1.ID#:** These 11 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. All 11 of these sequences cluster in phylogenetic analysis with subtype B sequences found in Thailand (sometimes referred to as B' or Thai-B). PCR products amplified from uncultured PBMCs were directly sequenced [Brown (1996)]. This report also included subtypes C and CRF01(AE) in Malaysia. Accession numbers U65538–U65548.
 - 119) **NL.114C** This sequence is one of many sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Samples were collected from the infant at birth, at 6 weeks

- and at 9 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included here [Mulder-Kampinga (1993), Mulder-Kampinga (1995)]. Infant 114 sequence is from Accession numbers L21111–L21153. Mother 114 sequence is from Accession numbers L21028–L21110. Infant 127 sequence is from Accession numbers Z47817–Z47832. Mother 127 sequence is from Accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in the entries with accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.
- 120) **NL.168:** This is sequence from one of 3 clones after culturing in PBMC. The isolate was originally from an AIDS patient in Amsterdam. [Wrin (1995)]. Accession numbers U15030–U15032. A V3-loop (105 bp) segment from the original isolate has been previously reported [Fouchier (1992)]. Accession number L06694. Neutralization assays were done with other clones of this virus [LaCasse (1998)] accession numbers AF035532–AF035534. LaCasse state that the primary isolate was SI phenotype and used both CCR5 and CXCR4 receptors and that a T-cell line adapted isolate derived from the primary isolate used only CXCR4.
- 121) **NL.H12178C:** This is a representative sequence from 6 patients infected with the same unit of blood in Amsterdam. [Wolfs(1992b)]. Accession numbers U04530–U04534.
- 122) **NL.X1:** This is a sequence from one of 10 clones from a recipient in a donor-recipient study. Sequences from donor Y and recipient X2 are also part of this study, but are not included here [Cornelissen (1995)]. Accession numbers Z47505–Z47514 are from X1. Other new sequences analyzed in this paper include Z47411–Z47540. Sequences M91828–M91838 (donor H and recipient O, referred to as patients A14 and A13, respectively in [Wolfs (1992a)] see B_NL1.A13) were also re-analyzed in this study.
- 123) **NL1.ID#:** These nine sequences are part of a study of presumed donor-recipient pairs from an HIV-1 transmission study conducted in the Netherlands. If pairs were extremely close or identical, only the recipient is included here. Recipient samples were from the first sample to be antibody positive, and are numbers 1,3,5,7,9, and 13. These sequences are from sequences of multiple clones from PCR amplified serum RNA [Wolfs (1992a)]. Recipient A1 was also studied as patient H1 in [Kuiken (1993)] and so is not included here (see B_NL4.H1). Accession numbers M91819–M91827, M91829, M91831–M91832, M91839, M91857–M91870, M91872, M91874, M91881–M91884, M91891, M91893, 91895–M91908, M91910, M91911–M91926. Number 13, and the donor were also analyzed in [Cornelissen (1995)].
- 124) **NL2.ID#:** These two sequences are part of a Dutch study of mutations occurring over a five year period (starting in 1985) in two patients. Serum RNA was PCR amplified and multiple clones were sequenced [Wolfs (1991)] and [Zwart (1992)]. Accession numbers M74591–M74684.
- 125) **NL3.ID#:** These six sequences from the Netherlands are samples from AIDS patients, using serum samples to generate PCR clones from viral RNA for sequencing [Zwart (1993)]. Accession numbers L01282–L01297.
- 126) **NL4.ID#:** These 74 sequences represent a study of early seroconverters from different times with different risk factors for transmission during the AIDS epidemic in the Netherlands. The year the sample was taken is indicated in the last part of the sequence name. The risk group of the individual from whom the virus is derived is indicated in the first letter of the sequence name (I, B and H for IVDUs, hemophiliacs, and homosexuals, respectively). Viral genomic RNA from sera was PCR amplified and amplification product was directly sequenced [Kuiken (1993)]. Accession numbers Z29219–Z29225, Z29256–Z29258, and Z29262–Z29325, U23670.
- 127) **NL5.ID#:** These 18 sequences are from a study of patients with, and without, AIDS dementia complex (ADC) in the Netherlands. Not all patients were originally from the Netherlands. Samples were collected between 1986 and 1992. Viral genomic RNA from sera and/or cerebral spinal fluid was reverse transcribed and PCR amplified and clones were sequenced [Kuiken (1995)] Accession numbers Z37531–Z37534, Z37734–Z37963, Z37970–Z37971
- 128) **NL6.ID#:** 16 is one of four sequences used in a study of HIV-1 envelope-mediated syncytium formation. Two clones were SI and two are NSI. 320 is a single SI clone. The sequences were derived from PCR amplified DNA from provirus cultured in PHA-stimulated PBMCs [Andeweg (1992)]. Accession numbers L08655–L08662. Complete genomes from patient 320 SI and NSI strains are found in entries with accession numbers U34603 and U34604 [Guillon (1995)], tat and other genes from patient 320, some

- of the same isolates, are found with Accession numbers M64489–M64492 [Groenink (1992)]. Another env sequence from patient 320 is found with accession number AF069524 [Follis (1998)].
- 129) **NL7.ID#:** These two sequences are from sets of sequences (Accession numbers U05797, U13240, U13241, U13243–U13247 for 537 and U13242, U13248–U13252 for 1058) used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. Both sequences were derived from PCR amplified PBMC DNA. Consensus 537 represents a set of sequences from a Dutch patient with a relatively stable CD4+ cell count at 62 months post-seroconversion. Consensus 1058 represents sequences from another Dutch patient whose CD4+ cell count at 73 months post-seroconversion was declining faster than 537's [Delwart (1994)]. See also US10.ID#.
- 130) **NL8.672** This is a sequence from a patient early in infection, before, or around the time of seroconversion. Three other patients studied in this paper (537, 1058 and 594), had previously been reported. See B_NL4.594, B_NL7.1058 and B_NL7.537 [Shpaer (1994)] and [Delwart (1995)]. Accession numbers U23651–U23663, U23667, U23670. See also US11.ID#.
- 131) **NL9.ID#:** These 19 sequences are from a cohort of homosexual men living in Amsterdam who seroconverted between 1985 and 1989. The sequences are from direct sequencing of PCR products after RT-PCR from serum RNA. Samples for h1, h139, h491, h1140 and h1234 were obtained at seroconversion. Samples for h138 and h1136 were obtained 12 months and 29 months after seroconversion respectively. The sequence for h320 was obtained from proviral DNA, 2 months after seroconversion [Zwart (1994a)] and [Zwart (1994b)]. Accession numbers L25884 and U05786–U05808. More sequences from this same cohort of men were published in [Kuiken (1996b)] and [Kuiken (1996a)] Accession numbers Z67875–Z67876, Z67885–Z67939, Z67941–Z67960, Z68015–Z68089, Z68109–Z68110. Some vpr, vpu and other region sequences are available from some of these patients as well. Some of the database entries for this set appear to be duplicates of sequences reported in other studies, for example the sequence from patient H39 (Z68061) is identical to the sequence with accession number U05787.
- 132) **NL10.ID#:** These 13 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. Accession numbers L76842–L76863, L76878, SR903853; L76897, SR9304737; L76879, SR911515; L76874, UM893272; L76912, UM9403860; L76888, SR925752; L76890, SR926969; L76894, GQ9301341; L76882, MA913670; L76886, SR923572; L76873, BR891413; L76910, UM9403051; L76885, GF921953.
- 133) **NL11.ID#:** These 2 sequences are from many sequences of 81 clones (patient N) and 105 clones (patient F) from serum, sigmoid tissue and fecal matter from each patient. All sequences from patient N were more similar to other sequences from patient N than to any other sequence in the database. Likewise all sequences from patient F were most similar to other patient F sequences. [van der Hoek (1996)]. Accession numbers Z76463–Z76648.
- 134) **NL12.ID#:** These sequences are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors found consistent differences in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication [Kuiken (1996b)]. See also B_GB5 and B_DE2 sets. Some of the patients in this study have been previously studied. For example, entries with accession numbers Z68061 and U05787 are both from the same patient and analyzed with the NL9 set.
- 135) **NL13.ID#:** These two sequences are from an Amsterdam Cohort Study on HIV infection and AIDS in homosexual men [van 't Wout (1998)]. ACH0208 and ACH0039 seroconverted during the course of the study and progressed to AIDS within 5 years thereafter. The subjects ACH0208 and ACH0039 had both SI and NSI virus variants. After seroconversion the SI variants were first detected then NSI in both the subjects. AF022285, AF021652 are the ones that are included here. The complete set has accession numbers AF022257–AF022302.
- 136) **NL14.ID#:** These 54 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were

- infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions.
- 137) **NL15.ID#:** These 7 sequences are from study of compartmentalization between intestine and blood. Serum and fecal samples were taken from 204 HIV-1 infected persons [van der Hoek (1998a)] and [van der Hoek (1998b)]. Of the 204 patients, 100 were diagnosed with AIDS-defining illness; 59 showed AIDS related symptoms that were not AIDS-defining; and 45 persons were asymptomatic. Of the 13 persons from whom clones were sequenced, 8 were diagnosed with AIDS-defining illness; 2 showed AIDS related symptoms that were not AIDS-defining; and 3 were asymptomatic. Results showed that in 6 persons the HIV-1 subpopulations in faeces and serum were similar, whereas in 7 persons distribution of V3 genotypes showed a marked difference. Database entries were created only for the 7 patients with a different distribution. Accession numbers AF034392–AF034408 and AF075711–AF075716.
 - 138) **NL16.ID#:** These 22 sequences are from study of patients from various risk groups in the Netherlands (Lukashov et al, unpublished 1998). Accession numbers AF071269–AF071290.
 - 139) **NL17.ID#:** These 5 sequences are from study of patients from the Netherlands, some of whom were undergoing suboptimal antiretroviral therapy [Nijhuis (1998)]. Many clones, from samples taken over time were sequenced for each patient. Only one of each is presented here. Accession numbers AF100001–AF100140.
 - 140) **NO1.ID#:** These 36 sequences are from Norwegian patients who were part of the Oslo HIV cohort study [Engelstad(1996)]. Uncultured PBMC DNA was PCR amplified in two nested PCR reaction steps. PCR products were directly sequenced. Where two peaks of equal height were observed at a single position, IUPAC ambiguity codes were used. Health, sex, year of sample (1989-1992), and risk group (IVDU, Het, Homo, Hemo) for each patient were noted in a table in the publication. Four sub sequences were also part of this set (X92913, X92914, X92917, X92918). Accession numbers for the 36 subtype B sequences are X92902–X92911, X92914, X92915, X92919–X92941.
 - 141) **NZ1.ID#:** These 8 subtype B sequences are from New Zealand [Dwyer (1998)]. Of the 10 isolates sequenced, 8 were subtype B and 2 were subtype C. Accession numbers AF052622–AF052629. Risk factors were heterosexual (NZ1, NZ6), homosexual (NZ2, NZ8–NZ12) and IVDU (NZ4). NZ1 and NZ6 were from females.
 - 142) **PR1.ID#:** These four sequences are from Puerto Rico, and were generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification of DNA from viral culture. The sequence ID numbers are abbreviated; for example D2PR732 can be read as DAIDS sequence (D), isolated in 1992 (2), Puerto Rico (PR), patient 301732 (732). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Accession numbers U04926–U04929.
 - 143) **PR2.ID#:** These 2 sequences are from Puerto Rico [Flores (1999)]. Pol protease region sequences were also presented in this study. Accession numbers AF096813, AF096814.
 - 144) **PY.ID#:** Ten sequences from 10 patients living in Asuncion, Paraguay. All 10 were male patients with symptoms of AIDS. Virus was propagated in tissue culture for an unstated length of time prior to harvesting proviral DNA for PCR and sequencing. PCR products were directly sequenced. PY.3614, PY.3615 and PY.12837 were syncytium-inducing and the other 7 were not. The tenth sequence, PY.3614p was PCR amplified directly from patient PBMCs with no culturing. The significant differences between the cultured sequence from this patient (PY.3614c) and the direct sequence, indicates that the virus that grew out in culture was a minority of the virus present in PBMC. The sequence of PY.3614c is not included in this alignment [Cabello (1995)]. Accession numbers U28949–U28959.
 - 145) **RE1.IDS#:** These 8 sequences are from Reunion Island [Bregel-Pesce (1999)]. Only subtype B which did not subtype well by HMA was sequenced. The majority of infections on Reunion Island were subtype B. Subtype A was also found. Accession numbers Y18591, Y18595–Y18601.
 - 146) **RU1.ID#:** These 2 sequences are from [Lukashov (1995)]. Accession numbers: L38405, RUS3A; L38407, RUS4A.
 - 147) **RU2.ID#:** These 15 sequences are from homosexuals living in St. Petersburg, Russia (Malykh et al, unpublished 1995). Accession numbers U40283–U40285 and U40319–U40330.
 - 148) **RU3.ID#:** These 13 sequences are from Rostov, St. Petersburg, Sochy, Ecaterinberg, Komi Republic, and Nizhny Tagil, Russia. Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. Although several of these cases were suspected to have epidemiological linkage

- (RU9 and RU10 are believed to have been infected by the same individual, sequences from the index case are not available; RU11, RU13 and RU27 were all thought to share a common index case, although none of them was directly infected by the index case), the sequences do not clearly indicate such epidemiological linkage and all 13 sequences are presented here [Leitner (1996b)]. Accession numbers U69646–U69652, U69654, U69660, U69662, U69666 and U69667.
- 149) **RU4.ID#:** These subtype B sequences are from a set from Russian IVDU and heterosexuals. The subtype B found among IVDU in southern Ukraine (for example UA1216, AF051512–AF051516) is similar in the env gene to the subtype B found in the CRF03-AB (KAL153) circulating recombinant form [Bobkov (1998a)] and [Liitsola (1998)]. Accession numbers AF051462–AF051465. Subtype B from homosexuals in Russia (RU1214, AF051474–AF051475) and in Ukraine (UA1229, AF051519–AF051520) were not significantly linked to the subtype B fueling the IV drug user epidemic. The sequence with accession number AF082457 is from an IV drug user, but is not tightly related to the more common IVDU B or AB recombinant sequences.
- 150) **SE.pt11s113:** This sequence is from one (patient 11 sample 113, collected in 1988) of a set of 13 samples from 9 epidemiologically linked individuals. The index case (patient 1) was a Swedish male who is believed to have contracted HIV while visiting Haiti in 1980. Six Swedish females were infected (patients 2,4,5,7,8 and 11) by patient 1. Two males (patients 6 and 10) were then infected by these females, and two HIV-infected children (patients 3 and 9) were born to the women. Sequences from each patient were determined by PCR amplification from uncultured PBMCs and direct sequencing. Heterogeneous sites were indicated with IUPAC codes. Extensive phylogenetic analysis was done to determine which methods accurately reconstructed the true phylogeny [Leitner (1996a)]. Accession numbers U68496–U68521.
- 151) **SE1.ID#:** These seven sequences are from a study of blood and CSF samples taken from each patient. The CDC disease stage class for the patients are as follows: II pts 930, 2815; III pt 931; IV-E pt 2951; IV-A pt 1032; and IV-C2 pts 1433, 1866 [Keys (1993)]. Accession numbers Z23178–Z23181, Z23185–Z23187, Z23192–Z23195, Z23200–Z23219, Z23224–Z23227, Z23232–Z23235, and Z23240–Z23255. Blood and CSF sequences from subtypes A and C were also studied.
- 152) **SE2.ID#:** These five sequences are from patients in Goeteborg, Sweden each of whom had recently seroconverted at the time of sampling in 1985-1990. Sequences from the sexual partners of all five patients who were believed to have transmitted the virus to these recipients were also determined, but are not shown here due to the great similarity to the recipient sequences (94% to 99% identity). Three of the recipients (R1, R3 and R4) were homosexual, and the other two were heterosexual. [Furuta (1994)]. Accession numbers U10929–U10950.
- 153) **SE3.ID#:** These 5 sequences are from a study of the ability of beta-chemokines RANTES, MIP-1alpha and MIP-1beta to inhibit primary isolates with SI and NSI phenotypes [Jansson (1996)]. The conclusion was that NSI viruses were inhibited by these beta-chemokines, whereas SI viruses were not. Each viral isolate was cocultured on PHA-stimulated donor PBMCs for 3 days, PBMC DNA was harvested, PCR amplified, and the PCR product directly sequenced. For each of these 5 patients (labelled patients A-E) one SI and one NSI isolate were sequenced, only one sequence is presented here. Accession numbers U76078–U76087.
- 154) **SE4.ID#:** These 4 sequences are from a study of 270 HIV-1 infected individuals living in Sweden [Alaeus (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 blood samples were available from 33 men and 42 women, from 15 different African countries. The 4 subtype B sequences were from individuals who were thought to have been infected in Kenya (SE8613 U76158), Eritrea (SE8875 U76167), Rwanda (SE7898 U76133) and Mauritania (SE7901 U76144).
- 155) **SE5.ID#:** These 2 subtype B sequences are from a study that includes the analysis of HIV-1 strains in seven cases of mother-to-child transmission in Sweden [Contag (1997)]. HIV-1 isolation from the mothers was conducted during the three trimesters, a few days after delivery and 6 months after parturition. The children were born between June 1989 and August 1992 and were followed by virus isolation at birth and at 3, 6, 12 and 18 months of age unless circumstances called for an extra sample. Breast-feeding was denied by all of the women described. See also A_SE.H4B, C_SE2.ID#, D_SE.H3 and AE_SE.H1. Accession numbers for subtype B are U56263–U56270, U56289–U56291.

- 156) **SE6.ID#:** These 10 subtype B sequences are from Sweden. Several clones were sequenced for each patient, sampled over time since primary infection took place [Karlsson (1999)]. The authors compared the diversity of sequences in primary infection with samples from subtypes B, C and CRF01(AE). Accession numbers for the entire set are AF014075–AF014104 and AF068484–AF068533.
- 157) **SG1.ID#:** These 26 sequences are from Singapore. Other sequences in this set were subtypes A, C and CRF01(AE), [Se-Thoe (1998)]. Accession numbers AF004235, AF004238, AF004239, AF004241–AF004246, AF004249, AF004250, AF004253–AF004256, AF004258, AF004260–AF004262, AF004264, AF004266, AF004267, AF004269, AF004270, AF004272 and AF004275
- 158) **SK1.ID#:** These nine sequences are from homosexual men living in the inner city of Bratislava, in the Slovak Republic. Patients 11, 12, 15 and 23 were classified as CDC stage A1 and were not taking any medication. Patients 18 and 20 were CDC stage B2 and were taking AZT. Patients 9 and 28 were CDC stage C3 and were taking AZT. No information is available for patient 22. Two other patients (10 and 51) were included in this study, but their sequences are not shown here because after their sequences proved to be similar to sequences from patients 9 and 28, respectively, epidemiological investigation indicated that they (9 and 10; 28 and 51) had been sexual partners. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in the databases. It is not stated whether the single sequence is a consensus of the eight clones or a single clone [Zachar (1996b)]. Accession numbers U53192–U53194, U53196–U53203.
- 159) **TH.92TH026:** This sequence is from Thailand. It is one of several env gene V3 region sequences obtained for the World Health Organization. Accession number U08717.
- 160) **TH.93TH067:** This sequence is from Thailand. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced [Penny (1996)]. Accession numbers U39258 and U39259.
- 161) **TH.T8174:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was an intravenous drug user and the sequence was obtained from PCR amplified PBMC DNA [Ou (1993)]. Accession numbers L19238 and L07446 are from the same patient. See also AE_TH.T8178.
- 162) **TH1.ID#:** These ten sequences are from individuals from Thailand. PCR-direct, peripheral blood PBMC DNA [Ou (1992b)] and [Ou (1993)]. (Published erratum appears in *Lancet* **342**:250 (1993).) Accession numbers L07442, L07449–L07456 and L07460.
- 163) **TH2.ID#:** The TB132 sequence is from a set of isolates from HIV seropositive individuals from Thailand. PCR, PBMC co-culture, DNA. Full env sequence is available [McCutchan (1992)]. Please note: the TB132 locus name in the database corresponds to the McCutchan et al. “BK132” isolate. Accession number L03697. The CM237 sequence is from PBMC proviral DNA [Mascola (1994)]. Accession number L14570. See also B_US14
- 164) **TH3.ID#:** These 2 sequences from Thailand are from asymptomatic individuals. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April 1994 supplement to the *Human Retroviruses and AIDS 1993 Compendium* [De Wolf (1994)]; [Osmanov (1994)]; [for HIV Isolation (1994)]. Accession numbers U08715–U08719, U08801–U08802, and U08783–U08784.
- 165) **TH4.ID#:** These twelve sequences are B subtype sequences from Thailand. Ten were genetically most similar to HIV-1 found in the Americas and Europe; these sequences were derived from people infected prior to 1988 (diagnosed in 1986 or 1987). The other two (N762 and N763) were designated B' and were isolated from people with more recent infections, 1988 and 1992. The sequences were obtained from PCR amplified PBMC DNA. The naming of the sequences submitted to the databases does not correspond with the naming of the sequences in the paper [Kalish (1994)]. Accession numbers for the entire set of thirteen sequences studied in this publication: U15576–U15588.
- 166) **TH5.ID#:** These three sequences are B subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and CRF01(AE) were analyzed. It is not clear from the paper or the database entries, which sequences came from individual 1 and which from 2 [Artenstein (1995)]. Accession numbers U21471, U21473, U21475. See also AE_TH6.ID#.

- 167) **TH6.ID#:** These 4 entries are from 1992 dried blood spot samples from Thailand. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from 3 previously identified HIV seropositive IV drug users (58, 11, 15) and a homosexual (86). Other samples from the same region were all CRF01(AE) (see CRF01_TH8.#). [Cassol (1996)]. Accession numbers U53310, U53311, U53314 and U53315.
- 168) **TH7.ID#:** These 70 sequences are from IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April, 1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced (except for patient 108, in which PCR product was cloned and 2 clones were sequenced, one shown here). Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were CRF01(AE). [Kalish (1995), Wasi (1995)]. Accession numbers U22543–U22547, U22549–U22552, U22554–U22556, U22558–U22560, U22562–U22566, U22568–U22574, U22576–U22603, U22605–U22608, U22610, U22613–U22616, U22618–U22623 and U22626. See also AE_TH9.
- 169) **TH8.ID#:** These 12 subtype B sequences are a subset of the 95 sequences reported in a paper out of which only 26 were submitted to the databases with accession numbers U85085–U85060 [Subbarao (1998)]. The samples of these sequences were collected from 215 asymptomatic HIV-1 individuals from June 1994 through January 1995 at 9 regional medical centers in northern, central and southern Thailand. Out of the 215 participants 65 were injecting drug users and 150 reported sexual risk behaviors out of which 51 were female sex workers, 41 attended antenatal clinics, 9 had STD's, 41 men with heterosexual behavior and 8 were men who had sex with men. Out of the 215 specimens subtyped 175 were subtype-E, 37 were subtype-B' and 2 were typical subtype-B. See also CRF01(AE) for the same study. Accession numbers for subtype B are U85064, U85066, U85070–U85075, U85079, and U85081–U85083.
- 170) **TH9.ID#:** These 6 subtype B sequences are from a study that was performed to evaluate the sensitivity and specificity of PEIA and HMA for determination of subtypes B and CRF01(AE) and to determine proportion of these subtypes in Bangkok over time [Wasi (1995)]. The paper and its related references give no significant information about the patients used in this study. This paper gives a brief account of only 3 of 9 patients: specimens 087 and 088 are both from 32 year old females who were infected in 1988 and specimen 091 is a 40 year old male who was infected in 1988. Accession numbers U22614, U22615, U22619, U22620, U22621 and U22622.
- 171) **TH10.ID#:** Accession numbers for the subtype B sequences are AF082553,
- 172) **TH11.ID#:** These 7 sequences are from Thailand (Subbarao et al, unpublished 1999). They seem to be significantly related to the RL42 sequence from China (accession U71182). The other 32 sequences in this study were of the CRF01(AE) form. Accession numbers for the subtype B entries are AF151767–AF151773.
- 173) **TT.QZ4589:** This sequence is from Trinidad (Blattner, W. et al, unpublished 1995). Accession number U32396.
- 174) **TW.TWCYS:** This sequence is from Taiwan (Huang, L. et al, unpublished 1998). Accession number AF086817.
- 175) **TW1.ID#:** These 16 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang (1997)]. Three subtype B sequences in this set were greater than 97% identical to the HXB2/LAI lab strain (see B_FR.LAI) of HIV-1, and are not included here (TW83, U73049; TW271, U73059; and TW335, U73061). They have been withdrawn from the databases. The manuscript reports that 123 of 143 sequences from Taiwan were subtype B, but only 27 of the 143 sequences were submitted to the sequence databases. Other subtypes found in Taiwan in this study were CRF01(AE) (17 cases), C (1 case), F (1 case) and G (1 case). Accession numbers for B subtype are U73045–U73054, U73056, U73057, U73059, U73061 and U73063–U73069.
- 176) **UA1.UA1216:** This sequence is from Kiev, Ukraine and is subtype B in both env and gag [Bobkov (1998a)]. Subtype A was also found in the Ukraine, see the UA entry in the subtype A section. This sequence, as well as another one from Ukraine (UA1229, no database entry submitted) and two from Russia (RU1229, AF051462–AF051465 and AF051466; RU1214, AF051474–AF051475) were subtype B in both gag and env genes, and are extremely similar in the env sequences to the env sequences found in the CRF03_AB circulating recombinant form (A in gag, B in env) which is common among IV drug users

- in Kaliningrad and other Russian and Ukrainian regions. Accession numbers for env are AF051512–AF051514 and for gag are AF051515, AF051516, AF051519, AF051420. The env gene sequences for the UA1229 sample have not yet had database entries created.
- 177) **UA2.ID#:** These 3 subtype B sequences are from an IV drug users in Nikolaev, Ukraine (Nabatov et al, unpublished 1998). Accession numbers AF100936–AF100938. A fourth sequence (NIK81, AF098952) which was found to be subtype B in env, is most likely A/B recombinant, based on its similarity to the other CRF03_AB sequences. It is listed separately with the CRF03 sequences.
- 178) **UA2.ID#:** These 4 subtype B sequences are from Ukraine (Grebjenjuk et al, unpublished 1998). Accession numbers Y16075, Y16078, Y16081, Y16082, Y16083. UA2 and UA4 were identical, so only UA2 is presented here. Y16079 is A/B recombinant, and Y16080 is subtype A.
- 179) **US.ACP1:** This virus was cultured from a seronegative man with Kaposi's sarcoma. (See: [Ho (1989)]). ACP1 was the sequenced after one passage in PBMCs. Accession number M80660. The sequence AC-H9 (M80661) was also derived from this patient [Ashkenazi (1991)].
- 180) **US.AD39:** This sequence is one recipient sequence from a set of 36 clones from a homosexual donor-recipient pair. Other donor-recipient pairs were studied by quantitative homoduplex tracking assay (QHTA) but only this pair was sequenced [Zhu (1996)]. Accession numbers U50780-U50815.
- 181) **US.ADA:** This sequence is from the monocytropic U.S. isolate ADA [Westervelt (1991)]. The complete genome of a derivative of this isolate has been reported to have been sequenced in [Theodore (1996)]. The complete genome is from a macrophage tropic HIV-AD8 isolate, derived from HIV-AD87, which was in turn derived from HIV-ADA. Accession numbers M60472, AF004394.
- 182) **US.ALA1:** This sequence is from an infectious clone of the 1985 U.S. isolate AL-1, taken from a patient with AIDS (Buckler-White et al unpublished 1988). Accession number M38430. AF133371, AF133376, AF133381, AR034235, I28919
- 183) **US.BAL1:** This sequence is from the macrophage tropic U.S. isolate BAL, harvested from lung alveolar tissue. Reitz M, et al., Unpublished (1990). Accession numbers M68893, M68894. The entry with accession number M63929 is 98% identical to this sequence and is also derived from the BAL isolate [Hwang (1991)]. The BAL1 sequence has all four characteristic signature sites of the "Dutch IDU" sequences of subtype B described by [Lukashov (1998b)]. The patient was from upstate New York, and his risk factor(s) are unknown; Marv Reitz, personal communication. The BAL isolate was one of the HIV-1 strains assayed for replication kinetics in macrophages taken from identical twins, and match unrelated donors [Naif (1999)]. See accession numbers AF133346, AF133351, AF133356, AF133361, AF133366, AF133371, AF133376 and AF133381. Some of the entries with accession numbers AF105559–AF105594 and AF105855 from patient D in Italy are nearly 96% identical to the BAL isolate, and may possibly represent PCR contamination events, but could also just reflect the European IVDU signatures in the BAL sequence.
- 184) **US.BCSG3:** This is a fragment of a full genomic sequence from the provirus SG3, cloned as a single proviral unit. This clone replicates more efficiently in chimpanzee than in human lymphocytes, and is extremely cytopathic and syncytium inducing (SI) in immortalized human T-cell lines. [Ghosh (1993)]. Accession number L02317.
- 185) **US.BRVA:** This sequence is from U.S. isolate BRVA, which was taken from the brain tissue of a AIDS patient with neurological disorders. [Anand (1989)]. Accession number M21098.
- 186) **US.BWB:** This sequence is from sequences derived from PCR amplified PBMC DNA from brain tissue [Monken (1995)]. Accession numbers L17088–L17126.
- 187) **US.CDC42:** This sequence is from an infectious clone of the U.S. isolate CDC-451 [Desai (1986)]. It was isolated in 1984 from a 16 year old hemophilia A patient, who died of AIDS in June, 1984. The viral isolate was propagated on H9 cells prior to cloning into Lambda phage and M13 phage for sequencing. Accession number M13137.
- 188) **US.DH12:** The DH12 isolate has been extensively characterized [Shibata (1995), Cho (1998)]. It is dual-tropic, using CCR3, CCR5 and CXCR4 coreceptors. Chimeric molecular clones with the macrophage-tropic AD8 isolate have been made, showing that either V1-V2 or V3 regions of env from DH12 can confer the ability to use CXCR4 onto AD8. The DH12 isolate was passaged in human and chimpanzee PBMCs prior to cloning. A complete genome sequence is available. GenBank accession numbers AF069139 and AF069140.

- 189) **US.Diaz:** This sequence is one of a set of 223 closely related sequences. All 223 sequences came from 3 patients with a common source of HIV, a blood donor and two recipients of this donor's blood [Zhang (1997)]. Accession numbers for all 223 sequences are U29433–U29437, U29956–U30145, and U43035–U43054. Gag sequences from 12 clones are also available in entries U31573–U31584.
- 190) **US.FASH:** This sequence is also known as 91US005.11 from the WHO Global program on AIDS. It is from a 17 year old female with primary symptomatic infection (PSI) from Birmingham, Alabama, USA. The risk factor for this individual was reported as heterosexual contact. The sequence indicates that this clone is subtype B. Blood was drawn in 1991. Although this clone has a 34 bp deletion followed by a premature stop codon in the envelope gene, relative to other subtype B sequences, the env protein was strongly positive in a CAT complementation assay. The patient from which this clone was isolated experienced rapid CD4 decline. In addition to the 17 amino acid truncation of the gp41 peptide, the env gene has a mutation at amino acid position 721 (bases 2157-2159) replacing tyrosine with cysteine (Y721C). This mutation has been shown in SIV to increase cell surface concentrations envelope glycoprotein [LaBranche,C.C. et al, J. Virol. 69: 5217- 5227 1995]. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao (1996a)]. Accession number U27434.
- 191) **US.HOBR:** This sequence is also known as 91US006.10 from the WHO Global program on AIDS. It is from a 28 year old male with primary symptomatic infection (PSI) from Birmingham, Alabama, USA. The risk factor for this individual was reported as homosexual contact. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao (1996a)]. Accession number U27443.
- 192) **US.JFL:** This sequence is from a non-infectious clone from the monocytropic U.S. isolate JFL [McNearney (1990)]. Accession number M31451. Other sequences from HIV-1 isolates epidemiologically linked to this isolate can be found in database entries with accession numbers L06256–L06273 [McNearney (1993)].
- 193) **US.JM:** This sequence, along with B_US.WM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. [Ashkenazi (1991)]. Accession number M80662, M80661.
- 194) **US.JRCSF:** This sequence is from an infectious clone of the 1986 U.S. isolate JRCSF, derived from from the CSF of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRFL was isolated from the brain of the same patient. JRFL does not replicate in Jurkat, U937 or HUT78 cells. JRFL does replicate in mononuclear phagocytes, and the macrophage-tropic region of the virus was determined by domain swapping with NL4-3 to reside in a 157 amino acid region of gp120 including the V3 loop [O'Brien (1990)]. Accession numbers U63632, M38429, U45960. Also see: [Pang (1990), Pang (1991), Koyanagi (1987), Klasse (1996)].
- 195) **US.MN:** This sequence is from U.S. isolate MN, taken from a 6 year old boy with AIDS from the Newark, New Jersey area in 1984. His mother was an IV drug user who died of pneumonia in 1982. His father was also HIV seropositive. A complete genome of isolate MN is found in M17449. Other sequences from this patient from the 1984 blood sample and from a 1987 sample taken shortly before death (U72495) are available also. [Reitz Jr. (1992)] [Gurgo (1988)]. Accession number M17449. See also L48364–L48379 [Lukashov & Goudsmit(1995)]. Another complete genome of isolate MN is found with accession number AF075719.
- 196) **US.NY5CG:** This sequence is from the 1984 U.S. isolate NY5. [Willey (1986)]. Accession number M38431. See also GenBank accession number K03346. A recombinant between NY5 and LAI has also been extensively studied, see B_FR.LAI entry.
- 197) **US.P896:** This sequence represents a molecular clone from an primary isolate derived from a Jamaican man who immigrated to Philadelphia 15 years earlier. At the time of viral isolation, he had no antiviral therapy, but was an AIDS patient with < 10 CD4 cells per mm³. The infectious molecular clone from which this sequence was derived is both macrophage-tropic and extremely cytopathic in lymphocytes [Collman (1992)] and [Kim (1995)]. Accession numbers M96155, U39362.
- 198) **US.R2:** This sequence is from the USA, from a donor of serum which contained a broadly neutralizing antibody [Quinnan (1999)]. Accession number AF128126.
- 199) **US.RF:** This sequence is from the full-length lambda clone HAT-3, from Haitian isolate RF. RF is from a 28 year old Haitian male who had moved to the United States at age 25, in 1980. He had no history

- of IV drug use, homosexuality or blood transfusions. In October 1983 he had 20 lb weight loss, giardia with diarrhea, thrush, and diffuse lymphadenopathy. His CD4/CD8 ratio was 0.08. He died in December, 1983. Primary culture from a November 1983 blood sample was co-cultured on HUT-78 cells [Reitz Jr. (1992)] [Starcich (1986)] [Popovic (1984)]. Accession numbers M17451 and M12508.
- 200) **US.RJS:** This is one of six biologically characterized clones from patient RJS, isolate 4. The HIV-1 infected individual had been infected for five years at the time of isolation in 1985. Patient RJS was a 31 year old homosexual male from California, who reported having sexual encounters with at least 1000 partners while HIV-infected, from 1980 to 1985 [Hahn (1986)]. Virus was isolated via coculture on donor PBMCs prior to cloning and sequencing. Complete env sequence is available [Daniels (1991)] and [Fisher (1988)]. Accession numbers M37491 and M37573–M37577.
- 201) **US.SB(A-C):** These three sequences are from 1988 U.S. isolates taken from a woman, her daughter and her sexual partner. The three viruses are epidemiologically linked, however the amino acids sequences appeared sufficiently divergent in this region to merit the inclusion of all three samples. The diversity is due to the fact that these family members had been infected for many years, prior to the 1988 sampling date. Sequences were directly sequenced from PCR amplification products after the virus was briefly cultured [Burger (1991)]. GenBank accession numbers M77228–M77230.
- 202) **US.SC:** This sequence is from the 1984 U.S. isolate SC, from an AIDS patient [Gurgo (1988)]. Accession number M17450.
- 203) **US.SF128:** This clone was isolated from the spinal cord tissue of a patient with dementia, after coculture with PBMCs. It is macrophage tropic, infecting macrophages but not T-cells [Liu (1990)]. Accession numbers M95292 and M38673. The ability to infect macrophages versus HUT 78 cells was mapped to the region between a *StuI* site in *env* and a *XhoI* site in *nef*, by replacing this region in SF2 with the same region from SF128.
- 204) **US.SF162:** This sequence is from an infectious clone from the U.S. isolate SF162, cultured from the cerebrospinal fluid of a patient with toxoplasmosis [Cheng-Mayer (1990)]. Accession numbers M38428, M65024.
- 205) **US.SF2:** This sequence is from an infectious clone from the U.S. isolate ARV-2. ARV-2/SF2 was isolated from a patient with oral candidiasis after co-culture with mitogen-stimulated PBMCs in 1984 [Levy (1984)]. [Sanchez-Pescador (1985)]. Accession numbers K02007 and I07977. HIVSF13 (Accession number L07422) is a more infectious virus taken from the same patient five months later, when he had developed Kaposi's sarcoma and *Pneumocystis carinii* pneumonia [Cheng-Mayer (1991)]. SF2 and SF13 are 98% identical to one another. The variation of SF2 in 9 years of infection in a chimpanzee has been studied by [Fultz(1997), Fultz (1997)], accession numbers U56884–U56887. This chimpanzee, infected by both SF2 and LAV-1b strains of HIV-1 was studied and developed AIDS [Novembre (1997), Davis (1998), Wei & Fultz(1998)] accession numbers AF006015–AF006032, AF027771–AF027785, AF049494 and AF049495.
- 206) **US.SF33:** This sequence is from an infectious clone from the 1984 U.S. isolate SF33 [York-Higgins (1990)]. Accession number M38427.
- 207) **US.twinAB:** This sequence is one of a set of 27 env sequences from a pair of heterozygotic perinatally HIV-1 infected twins who were observed during their first 2 years of life. Twin A remained asymptomatic through her first 2 years while twin B developed AIDS at six months and died at 22 months of age. Patient PBMCs were cocultured with donor PBMCs, prior to DNA extraction and PCR amplification. Approximately 500 copies of HIV proviral DNA per PCR reaction were used. PCR products were cloned prior to sequencing. Viral phenotypes from both infants, and all time points were also assessed. All were found to be non-syncytium inducing, but they differed in their ability to infect primary macrophages. Overall, it seems that the production of neutralizing antibodies by the healthy twin was the most important clinical factor [Hutto (1996)]. GenBank accession numbers U47562–U47588 for the envelope sequences and U47589–U47613 for *tat* sequences from the same blood samples.
- 208) **US.UNC116:** This subtype B sequence is from a patient UNC116 who was born in 1969 with severe Haemophilia A and received over 50000 U of non-heat treated factor VIII concentrates between 1978–1984 [Michael (1998)]. The parents of the subject were CCR5^{-/-} and CCR5^{+/-}. The samples were obtained from patient UNC116 between 1985 and 1992 were of SI genotype based on presence of positively charged amino acids in the V3 loop region. UNC116 is the first demonstration of exclusive and persistent CXCR4 usage in an HIV-1 infected individual. Accession number AF034385.

- 209) **US.WCIPR:** This subtype B sequence is from the USA (Fang et al, unpublished 1999). A complete genome has been sequenced. Accession numbers U69584–U69593 and AF003887.
- 210) **US.WEAU:** This sequence is from a cultured isolate from a patient described as “patient 1” in [Clark (1991)] and as WEAU 0575 in [Piatak Jr. (1993)]. The sequence is from a fully infectious complete, cloned genome. It has a high tropism for T-cells and is syncytium-inducing. Accession number U21135.
- 211) **US.WM:** This sequence, along with B_US.JM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals [Ashkenazi (1991)]. Accession number M80663.
- 212) **US.WMJ22:** This sequence is from the isolate WMJ22, isolated from a 4 year old female with AIDS, born in Florida to a woman of Haitian descent living in the U.S. Virus was cocultured on an immortalized T-cell line [Hahn (1986)] and [Starcich (1986)]. Accession numbers M12507, K03457.
- 213) **US.WR27:** This sequence is from a complete genome. It represents the first complete PCR-derived sequence of a U.S. clinical isolate of genotype B expanded only in primary PBMC. This provirus harbors a uniquely truncated V3 loop. WR27 was a patient with clinical progression to WR stage 5 when blood was drawn for viral isolation in 1988. This sequence is a from a PCR clone from a primary isolate that was expanded in PBMC. The virus isolate had an SI phenotype [Salminen (1995)]. Accession number U26546.
- 214) **US.YU:** This is one of eight lambda phage clones and 12 PCR amplified clones derived from the uncultured brain tissue of a patient with AIDS dementia complex. A macrophage tropic clone (YU-2) is almost identical to the consensus sequence of YU in this region, with only a single amino acid change (K to N) 10 amino acids from the carboxy-terminal end of the sequence [Li (1991)]. Complete genomic sequences are available for two of the HIV1 YU clones, along with biological characterizations of four of the HIV1 YU clones: [Li (1992)]. The GenBank accession numbers for the YU-10 and YU-2 complete genomes are M93259 and M93258, respectively. Accession numbers for the other clones are M89972–M89984.
- 215) **US1.ID#:** These are forty 1990–1991 U.S. samples, from the study of the dentist who was thought to have been the source of HIV-1 infection of six of his patients. Only the dentist’s viral sequence and the Florida control sequences are shown here; the six epidemiologically and genetically linked patients are excluded from this alignment because their viral sequences were very similar to the dentist’s. All sequences were PCR amplified from patient PBMCs. Most are direct sequences from the amplification products, although some are sequences from one of multiple clones of PCR products [Ou (1992a)]. Accession numbers for the 75 sequences in this set: M90847–M90853, M90881–M90886, M90894–M90900, M90907–M90912, M90914–M90956, M90958–M90964, M92100–M92133, L22590–L22606 and U06872–U06919. See also [Korber & Myers(1992), Crandall(1995), Smith & Waterman(1992), DeBry (1993), Palca(1992b), Palca(1992a), Abele & DeBry(1992), Hillis & Huelsenbeck(1994), Ciesielski (1991), Ciesielski (1994)]. In a related study [Delwart (1995)] heteroduplex mobility tracking assays were used to re-analyze samples from this set, again supporting the dental to patient transmission hypothesis.
- 216) **US2.ID#:** These 15 sequences are from the USA, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification products of DNA from viral culture. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Full length envelope genes from some of these clones have been expressed, [Gao (1996a)]. Accession numbers U04907–U04915, U04918 and U04921–U04925.
- 217) **US3.ID#:** These four sequences are from the US, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. gp160 sequences of clones from expanded culture stocks are available. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Accession numbers U08448–U08449 and U08451–U08452. Accession numbers for additional clones from these patients: U04916–U04917 and U04919–U04920.
- 218) **US4.ID#:** These sequences are from five primary seroconverters. PCR-clones, peripheral blood PBMC DNA. Although sequences were also available from two of the donors, only sequences from five recipients are shown here [Zhu (1993)]. Accession numbers L21224–L21328, L21331–L21348,

- L21372–L21373, L21376–L21377, L21380–L21381, L21384–L21389, L21393–L21394, L21397–L21398, L21400–L21401, L21404, L21405, L21408, L21410–L21411, L21414, L21417–L21418, L21421, L21426–L21427, L21430–L21432, L21434–L21437, L21440–L21442, L21445, L21447, L21449–L21450, L21453–L21454, L21457, L21459, L21462–L21463, L21468, L21472–L21515, L21519–L21520, L21522–L21523, L21525–L21526, L21528–L21533, L21535–L21536, L21538–L21539, L21541, L21543, L21545–L21546, L21548, L21550–L21551, L21553–L21554, L21556, L21558–L21559, L21561–L21569, L21571–L21578, L21580, L21582, L21584, L21586, L21588 and L21590.
- 219) **US5.ID#:** These six sequences were part of a study of blood sequences compared to brain sequences from six individuals [Korber (1994a)]. Accession numbers U05360–U05568.
- 220) **US6.ID#:** These 7 sequences are each one of several cloned PCR products from PBMC proviral DNA from an individual infant who was part of a study of mother-infant transmission [Ahmad (1995)]. Infant blood samples were taken from 1 week (infant 7) to 34 months (infant 4) post-partum. Dates of sample collection were: Infant 1, 10/25/91; Infants 2–4, 10/31/91; Infant 5, 2/6/92; Infant 6, 8/30/93; Infant 7 5/13/93. Maternal sequences were also reported as part of this set. Accession numbers U16390–U16652.
- 221) **US7.ID#:** These 3 sequences are each one of several cloned PCR products from PBMC proviral DNA from an individual patient. These sequences were part of a study of early samples (1984–1986) from the San Francisco region. One of the samples (552-3) was HIV negative but was determined to be contaminated with blood from another (565-3) so the six samples from 552-3 and 565-3 are represented by one. Another sample (552-5) from patient 552 was not contaminated, and it is presented as B_US11.552, because a larger region was reported in that publication [Sabino (1994b)] Accession numbers L20371–L20380. More V3 and tat sequences from these individuals are discussed in [Sabino (1994a)]. Accession numbers U00243–U00399 and U01513–U01529.
- 222) **US8.ID#:** These 2 sequences were from a study of three recipients of contaminated blood. Recipient 1 (R1) and recipient 2 (R2) each received blood from different donors. A third recipient, not presented here, received blood from both donors. All three recipients were neonates. R1 received erythrocytes from donor 1 on 19 October, 1984 at the age of 3.5 weeks. For R1 the sequence of one of the two clones (2E) is presented here. R2 received erythrocytes from donor 2 on 24 September, 1984 at the age of 2 months. For R2, one of 6 clones is presented here. Blood samples for this study were drawn in March 1986. R1 had slow weight gain, and R2 had lymphadenopathy at time of sample collection [Diaz (1995)]. Accession numbers U11188 = R1–2E, U11189 not used; U11203, U11196, U11199, U11192–U11194 = R2 six clones. The tat and envelope V4–V5 regions of clones from these same individuals are also available in U11173–U11178, U11180, U11205–U11209.
- 223) **US9.ID#:** These four sequences are from several samples taken over a range of time from four different subjects. Blood samples for S1 were drawn in Nov '85, Jul '87, Jan '88 and May '89. Blood samples for S2 were drawn in May '85, Apr '87 and Oct '87. Blood samples for S3 were drawn in Jun '87 and Dec '87. Blood samples for S4 were drawn in Jan '85, Jan '89 and Jun '89. S2, S3 and S4 had decreasing CD4 counts during the study period. S1 had fluctuating CD4 counts. [McNearney (1992)]. GenBank accession numbers L03430–L03453 and L23575–L23588 = S1; L03454–L03477 and L23618–L23633 = S2; L03478–L03490 and L23589–L23600 = S3; L03491–L03515 and L23601–L23617 = S4.
- 224) **US10.ID#:** These three sequences are from sets of sequences used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. All sequences were derived from PCR amplified PBMC DNA. The MA145 sequence is one of the sequences (GenBank accession numbers U00821, U00822, U00831–U00839) taken from an asymptomatic male from Massachusetts over a period of 4.5 years starting April 1989. Patient MA, from the US, was infected in 1984 or 1985, and had been experiencing neurological disorders prior to 1989 [Kusumi (1992)]. The SFBU and SFPE sequences (Accession numbers U13373–U13380 and U13381–U13388 respectively) were taken from two patients with AIDS from San Francisco [Delwart (1994)]. Other sequences from patient MA can be found in database entries with accession numbers U00804–U00822, U00831–U00850, U00873–U00888, M79342–M79354 and M90025–M90046.
- 225) **US11.ID#:** These 11 sequences came from patients early in infection, before, or around the time of seroconversion. Many sequences for 306, 419, 349, and 074 are available [Shpaer (1994)] and [Delwart (1995)]. Accession numbers U23662–U23666, U23668, U23669, U23671–U23708, L20381. See also B_NL8.ID#.

- 226) **US12.ID#:** These six sequences were used in an investigation into the transmission of HIV-1 from one child (CHA), who had received zidovudine, to another child (CHB), who harbored a zidovudine-resistant strain. The presence of the zidovudine-resistant strain in child A and B, and the lack of such a strain in child 2's mother was used to show that child B was infected by child A and not by child B's mother. LC sequences are from children used as local controls. All sequences were derived from PCR amplified PBMC DNA [Fitzgibbon (1993)]. Accession numbers L12751–L12756, L19695, L19697, S66942. L12756 is listed as “isolate 100” in the databases, but seems to be the “group B consensus sequence” used for phylogenetic analysis.
- 227) **US13.ID#:** These three sequences are from three IV drug users in Florida. Proviral DNA sequences were obtained from blood, cerebrospinal fluid and dorsal root ganglia from each of the three individuals. Sequences for V1-V5 of env, were PCR amplified, cloned and sequenced [Shapshak (1995)] [Xin (1995b)]. The sequence for patient 149 is one of 24 clones with Accession numbers U16094–U16117. The sequence for 141 is from one of the database entries labelled as being from patient 141 with the exception of: R5D, R6D, R7D, R2D and R4D, which were similar to IIIB strains of HIV-1; and R1R, R3R, R7R, R8R and R9R, which were similar to samples from patient 144. The sequence for 144 is from one of the database entries labelled as being from patient 144 with the exception of: R3R, R6R, R9R, R12R, R13R and L1D, which were similar to IIIB strains of HIV-1; and C3D, C4D, C7D, C8D and C10D, which were similar to samples from patient 141. While infection of each individual with multiple strains of HIV (including one very similar to the IIIB lab strain) is a possible explanation of these findings, we are only including one sequence from each patient for this alignment. The authors are currently (1996) resequencing new samples from these patients. Accession numbers for patients 141 and 144 are U16032–U16093, U25191–U25261. Further analysis of these same patients was done later [Shapshak (1999)]. Accession numbers AF125810–AF125860.
- 228) **US14.ID#:** These four sequences are from DNA from PBMC [Mascola (1994)]. Accession numbers L14573–L14576. See also B_TH2, AE_TH2.
- 229) **US15.ID#:** These six sequences are from a study of infants. Blood samples were collected from six infants over time. [Strunnikova (1995)]. Accession numbers U22682–U22810, U22834 and U22835.
- 230) **US16.ID#:** These two sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. Patient ARTC1 was an asymptomatic individual from New York and ARTC3 was an AIDS patient from New York. [Pestano (1995)]. Accession numbers U11586–U11594. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.
- 231) **US17.ID#:** CB7 is one sequence from a set of 12 clones from two different samples (six clones each) from the same patient, collected in 1988 and 1990, plus two more clones as yet unpublished. The patient seroconverted in 1985. The patient did not receive any antiviral therapy until 1992. The patient's CD4 count was 1035 in 1988 and 807 in 1990. The patient's PBMS were cocultured with donor PBMC for an unspecified length of time before cultured DNA was isolated and PCR amplified. Individual clones of PCR product were then sequenced [Wang (1995)]. Accession numbers U16324–U16335 and U19706–U19711. The other 8 sequences are individual clones from 8 different patients, all from the same clinic in Boston, MA. Individual CB7 was again included in this study, [Wang (1996b)] [Wang (1996a)]. Accession numbers U60152–U60162 and U27658–U27669.
- 232) **US18.ID#:** These 22 sequence are from a study of a HIV-1 infected dentist and many of his HIV-1 infected patients. In this case, no evidence was found for dentist to patient, patient to dentist, or patient to patient transmission [Jaffe (1994)]. Accession numbers U11454–U11490. Unfortunately the dentist sequences were not submitted to the databases.
- 233) **US19.ID#:** These two sequences are from a comparison of a mother who transmitted HIV to her infant and a mother who did not. One of the sequences (nontransmitter-283 accession number U07839) was 99.4% identical to the LAI strain of HIV-1 and is not included here. Sequences from the transmitter and nontransmitter were highly similar with the exception of sequence nontransmitter-217 (U07836). Thus this sequence is included here as N217, and one of the others is presented as N210 [Ayyavoo (1996)]. Accession numbers U07833–U07841 and U07891–U07917.
- 234) **US20.ID#:** These three sequences are from a study of mother-infant transmission. Only child sequences are used here, but mother and child clonal sequences are reported in [Roth (1996)]. Proviral DNA was PCR amplified from uncultured patient PBMCs. Cloned PCR products were sequenced. Child 1

- was 27 months old, child 2 was 5 months old and child 3 was 21 months old when blood was drawn for sequencing. Sequences, including those from the mothers, are in database entries with accession numbers U47745–U47807.
- 235) **US21.ID#:** These six sequences are from six different patients, all men from the Chicago, Illinois, USA MACS cohort. In the study, P1 & P2 were rapid progressors, P3 & P4 normal progressors, and P5 & P6 non-progressors. Their estimated years of seroconversion are as follows: P1, 1985; P2, 1985; P3, 1986; P4, 1986; P5, 1985; and P6, 1984. A total of 292 sequences of the C2-V5 region of envelope were completed. In the database entries the sample numbers are of the form Px.y-z, where x stands for the patient number, y is the number of months after the estimated date of seroconversion, and z is the clone number. Each of the six sequences presented here is from one of the many clones from that patient [Wolinsky (1996)]. Accession numbers U35895–U36185.
- 236) **US22.ID#:** These 12 sequences are from a study of variability of tat and env genes [Lorenzo (1996)]. Accession numbers U57104–U57216. Tat sequences from these clones are also available (U57217–U57304).
- 237) **US23.ID#:** These 3 sequences are from a study of variability of vif and env genes [Sova (1995)]. Uncultured or short-term cultured PBMC proviral DNA was PCR amplified, and several clones from each PCR reaction were sequenced. Only one clone sequence from each patient is presented here. Accession numbers U50615–U50628. Vif sequences from these patients are also available (U41055, U41056, U41179–U41182 and U42229–U42282).
- 238) **US24.ID#:** These 4 sequences are from unpublished database entries by Schwartz et al. There seems to be either sequence mislabeling or interpatient sample contamination in this large set of sequences, because some sequences labelled as one patient (10185W1G for example) cluster with sequences from another patient (10187 in this case). Accession numbers U49518–U49640, U51311–U51326, U45860–U45876 and U49576–U49608.
- 239) **US25.RM:** This sequence is from a blood recipient M who received a red blood cell pack from donor W in October 1984. Blood was sampled from recipient M for this sequence in January 1986, 15 months after the transmission event. Sequences from donor W (collected in January 1987, 27 months after transmission) and another recipient (O, collected in November 1986, 20 months after an RBC pack transfusion in March 1985) are also available [Delwart (1995)]. Accession numbers U22826–U22828.
- 240) **US26.ID#:** These two sequences are from: A) a slow progressor, patient A, a 26 year old caucasian male who lost an average of 31 CD4+ cells/ml per year during the first 3 years of infection, and remains healthy; and B) a rapid progressor, patient B, a 38 year-old caucasian male who lost an average of 175 cells/ml per year during the first 3 years of infection. Both patients were infected from the same source but patient B picked up two divergent subtype B forms (one shared with patient A, and one unique to patient B). In the database entries, patient B is labelled as “14” and patient A as “13” [Liu (1997)]. Sequences were determined by PCR from uncultured PBMC DNA and RT-PCR from plasma viral RNA. PCR products were cloned prior to sequencing. One representative clone from patient A was picked for the “13A” sequence presented here, and one of the divergent B sequences was picked for the “14B” sequence presented here. Accession numbers U56146–U56235, U79034–U79113.
- 241) **US27.ID#:** These 6 sequences are from infants infected perinatally and followed over time after birth as part of the Los Angeles Perinatal Transmission Study [Ganeshan (1997)]. DNA sequences were determined from cloned PCR products from uncultured patient PBMCs.
- 242) **US28.ID#:** These 14 sequences are from studies of long-term survivors. Patient 3799 was transfused with HIV-1 infected blood in October, 1982 at the age of 30, and has remained asymptomatic, with very low viral load and fairly stable CD4 counts for over 13 years. The donor, as well as two other recipients of his blood, have all died from complications of AIDS. The husband of 3799, as well as two children born and breastfed after the 1982 transfusion, are all HIV-negative [Michael (1995)] and [Schwartz (1997)]. The other patients showed varying rates of disease progression. During the study, only 101867 (U45860–U45876, see B_US24 set) died of AIDS. Sequences were PCR amplified directly from patient PBMC without coculture, cloned and sequenced. Only one clone from each patient is presented here. Many of these sequence entries were either mislabelled in submissions to GenBank, or represent inter-patient sample contamination. For example the entry with accession number U51317 was

- labelled as 10188, but is related to the other sequences from patient 1873. Accession numbers U60670–U60733. Entries with accession numbers U24443–U24487 are also from patient 3799, but not V3 region. AF004025–AF004027 are patient 10189G.
- 243) **US29.ME1:** This sequence is from a 40 year old patient in the Pittsburgh AIDS Clinical Trials Unit [Chen (1997)]. The paper describes a new method, called “progressive amplification” for obtaining full-length infectious molecular clones of HIV. In this study, two clones were generated and the env gp120 completely sequenced. One (ME1) was from the patient when asymptomatic, and both the viral isolate (367), and the molecular clone (ME1) grew on macrophages but not T-cells (Macrophage tropic) and did not show cytopathic effect on either MT-2 or PBMCs (NSI). The other (ME46) was from an isolate (828) taken from the same patient 20 months later, when the patient had AIDS. This isolate and the clone derived from it were able to grow on both macrophages and T-cells. It induced cytopathic effect in MT-2 cells and PBMCs (SI). Accession numbers U66221, U66222. The paper also reports on sequences from “patient P” in figure 5b, but those sequences are not available.
- 244) **US30.ID#:** These 10 sequences are from rapid and slow progressors who were studied in 1990-1994. The subjects were all asymptomatic and taking AZT (zidovudine) at the beginning of the study. Patients A, C, E, G and I were classified as slow progressors, while patients B, D, F, H and J were classified as rapid progressors. Two sequences appear to be mislabelled: U69410 labelled patient G clusters with patient E, and U69380 labelled as patient E clusters with patient G. Eight of the ten patient F, 1993 sequences are suspiciously similar (greater than 98% identical) to the HXB2R lab strain of LAI/IIIB. Each sequence is a single clone from one patient, although several clones from each patient at two time points are available [McDonald (1997)]. Accession numbers U69282–U69481.
- 245) **US31.ID#:** This study demonstrated the influence of three monoclonal antibodies IgG1b12, 2G12, and 2F5 to the HIV-1 envelope glycoprotein, and a tetrameric CD4-IgG molecule (Cd4-IgG2), for the ability to neutralize primary HIV-1 isolates from the genetic clades A through F and from group O. Each of the reagents broadly and potently neutralized B clade isolates [Trkola (1996a)]. Accession numbers are U79721, U79720, U79719.
- 246) **US32.ID#:** These 29 subtype B sequences are from Hawaii (Hoffman, P.R. et al, unpublished 1998). Patients ranged in age from 27-48. Most of them were Caucasians. Risk groups included heterosexuals and IDU's. Accession numbers AF016547–AF016579.
- 247) **US33.ID#:** These 15 sequences are from patients from USA who were sampled many times, with several clones sequenced at each time point [Markham (1998)]. For each patient the first clone from the first sample is presented here. The complete set had accession numbers AF016760–AF016825 and AF089109–AF089708.
- 248) **US34.ID#:** These 14 subtype B sequences are assumed to be from the USA. Unpublished by deOliveira et al. One sequence in this set, with accession number U52058, was subtype A/G recombinant and similar to the IbNG circulating recombinant form. Accession numbers for subtype B are U51942, U52055–U52057, U52059–U52063 and AF022243–AF022246
- 249) **US35.ID#:** These 16 sequences are from San Francisco Men's Health Study participants who seroconverted while under observation. All patients had an envelope with subtype B. Eight percent of these individuals reported homosexual/bisexual contact, 20% were exclusively heterosexual not necessarily at risk [McCutchan (1998)]. Accession numbers AF025749–AF025764.
- 250) **US36.SC14:** This subtype B sequence is from a recent seroconverter (SC14) from the San Francisco Men's Health Study participants (McCutchan et al, unpublished 1998). Two sequences from this individual were hypermutated, and two were more normal. Analysis of this patient was also included in [McCutchan (1998)]. Accession numbers U90932–U90935.
- 251) **US37.ID#:** This subtype B sequence is from a study in which peripheral blood mononuclear cells were isolated by density gradient centrifugation from 2 HIV-1 seropositive volunteers who were participating in a study of cell mediated immune responses to HIV infection [Ray (1998)]. Initially neither of them had developed symptomatic HIV-1 infection, subject AA developed Herpes zoster ophthalmicus and died during the study period, subject BB(U78831) remained asymptomatic. Sequences from subject AA are not included here because they were too short, the complete gp120 of subject BB was available. Accession number L78831 for BB and L78832 for AA.
- 252) **US38.ID#:** These 2 subtype B sequences are from the donors in a study that describes a case of simultaneous transfusion of 2 HIV-1 infected units of blood into one individual [Diaz (1995), Diaz (1996)].

- The dual recipient patient was a 54-year male with oatcell carcinoma of the lung, being treated with prednisone at the time of the index transfusion. He was transfused in November 1984 with a pool of platelet concentrates, two of which were subsequently determined to have been from HIV-1 seropositive donors. Accession numbers for D1 are U43988, U43993 and U43994. Accession numbers for D2 are U43997, U44003–U44007 and U44010. The study also includes a set of tat sequences with accession numbers U43986–U43992, U43995, U43996, U43998–U44002, U44008, U44009, U44011–U44023. Dual recipient sequences are also available, see accession numbers U11136–U11189.
- 253) **US39.ID#:** These 49 subtype B sequences are from vaccine clinical trials [Connor (1998)]. Nine are from vaccinees who received a gp120 vaccine derived from MN. Eight are from vaccinees who received a gp120 vaccine derived from SF2. Two are from vaccinees who received placebo. The rest are from local controls who were not vaccinated. These were volunteers involved in clinical trials of MN and SF2 gp120 vaccines who became HIV-infected after vaccination. Accession numbers for the entire set are U84792–U84887.
- 254) **US40.ID#:** These 20 subtype B sequences are the result of study done from 1992–1994 in which a few newly infected HIV-1 patients who resided in South Bronx New York were studied [Irwin (1997)]. Out of these sequences 2 were subtype A and 20 were subtype B. Accession numbers U90181–U90200.
- 255) **US41.ID#:** These 5 sequences are from a study that included 106 sequences with accession numbers U96502–U96608. Only one clone from each patient is included here. One gag sequence with accession number Z97081, is also presented in this study [Delwart (1998)]. Semen and blood specimens were obtained from patients at Stanford University Medical Center. Patient JO and PE were infected for unknown length of time and each had clinical AIDS at the time of sampling. The PE sample patient died within one year of sampling while patient JO was asymptomatic on combination retroviral therapy 6 years after sampling and had a few kaposi's sarcoma lesions. Patients 613, 064 and MA were all asymptomatic and were infected for 1 yrs, 5 yrs and 6 yrs respectively prior to sampling. Accession numbers U96502, U96523, U96546, U96571 and U96579.
(Salzman et al, unpublished 1998). Accession numbers U81888–U81956.
- 256) **US43.ID#:** This sequence is one of 40 sequences taken from a child from Baltimore Maryland, USA who was infected perinatally [Strunnikova (1998)]. Accession numbers AF032926–AF032965
- 257) **US44.ID#:** This set of sequences contained numerous interpatient PCR contamination events as well as contaminations with the LAI, RF and MN lab strains of HIV-1 [Frenkel (1998)]. None of the sequences are used in the V3 analysis, because of the possibility of PCR-induced recombination. The publication pointed out that neonates who test positive for HIV-1 by PCR, and later revert to HIV negative, may be the result of PCR contamination events, rather than actual clearance of the virus. Accession numbers AF065488–AF065598.
- 258) **US45.ID#:** These 4 sequences are from the USA [Ostrowski (1998)]. Many samples were collected from each patient over time and many clones from each sample were sequenced. The study examined the effect of perturbing the human immune system with vaccination with tetanus toxoid and found that it temporarily increased viral load following the activation of T-cells. NSI virus was preferentially increased over SI virus in the one patient where both tropisms were found. Accession numbers AF080698–AF081019 and AF098022–AF098061.
- 259) **US46.ID#:** These sequences are from rural areas of the state of Georgia, USA [Roth (1999)]. Accession numbers AF086975–AF087001.
- 260) **US47.ID#:** These 6 sequences are from patients undergoing combination antiretroviral therapy. Many sequences were collected from each of 8 patients over time, to assess the rate of evolution and thus help determine the replication rate of virus in patients with low or undetectable viral load [Zhang (1999)]. Accession numbers AF093912–AF094325, AF130388, AF130389.
- 261) **US48.ID#:** These 12 sequences are from the USA [Voulgaropoulou (1999)]. Patient WL had two very diverse viral quasispecies, perhaps indicating dual infection, but only one of the 2 is presented here. Accession numbers AF094977–AF095123.
- 262) **US49.ID#:** These 14 sequences are from the USA [Zachar (1999)]. The study sequenced virus from mother's blood and from trophoblasts and cord blood of the infants. Accession numbers AF150060–AF150079.
- 263) **US50.ID#:** These eight sequences are from asymptomatic individuals identified after donating blood in Memphis, Tennessee, USA. [Slobod (1994)]. Accession numbers U09140–U09175.

- 264) **UY1.ID#:** These 9 sequences are from 22-37 year old patients with CDC stage IV disease (except 726 at stage III). All are from Montevideo Uruguay. Four were homosexuals (270 352, 093 and 672), two were IV drug users (406 and 726), and the other 3 were heterosexuals (1699, 1193 and 376) [D. (1996)]. Accession numbers U66414–U66422.
- 265) **VE1.ID#:** These 8 sequences are from 8 individuals in Venezuela. Patient PBMCs were cocultured with donor PBMCs. Proviral DNA was harvested PCR-amplified. PCR products were directly sequenced. Nearly complete env gp120 sequences were determined, as well as pol gene sequences. [Quinones-Mateu (1995)]. Accession numbers U16764–U16778, even numbers are env, odd numbers are pol.
- 266) **VN1.HCM9:** This sequence is from South Vietnam [Menu (1996)]. The sequence is from Ho Chi Minh city, from a woman infected by her HIV seropositive sexual partner who was thought to have been infected while traveling in Europe. Three other sequences in this study were found to be CRF01(AE). Accession number U29209. Another sequence from this same patient was reported in 1999 [Kato (1999)]. In this paper the authors state that the woman's partner did not live in Vietnam, and the patient is called HCM312, but the sequences are more than 97% identical to each other, and the authors agree that they both sampled the same patient. Accession number AB025098.
- 267) **ZA.0117:** This sequence is from a study of 72 seropositive women from South Africa [Moodley (1998)]. The mean age was 26 years. Patient 0117 was asymptomatic. Data from this study shows the dramatic growth of HIV-1 subtype-C in this population in South Africa. See also C_ZA3.ID# and A_ZA.134. Accession number AF053279.
- 268) **ZA.009:** This subtype B sequence is from South African sequences published in [Van Harmelen (1999)]. Accession number AF095828. Subtypes A, C and D were also found in this study in South Africa.
- 269) **ZA1.ID#:** These 7 sequences are from 7 individuals in South Africa. ZA504 was from a 33 year old white male homosexual with AIDS and the virus was syncytium-inducing (SI). ZA508 was from a 32 year old white male bisexual with ARC and the virus was NSI. ZA509 was from a 30 year old white male homosexual with AIDS and the virus was SI. ZA524 was from a 49 year old white male bisexual with AIDS and the virus was NSI. ZA510 was from a 29 year old white male heterosexual with ARC and the virus was SI. ZA512 was from a 26 year old white male homosexual with ARC and the virus was SI. ZA513 was from a 3 year old black male blood transfusion recipient with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product [Engelbrecht (1995)]. Accession numbers L48063–L48066, L48069, L48071 and L48073. Database entries U33770 and U33774–U33779 are shorter env gene fragments from these same clones.

C Subtype

At this time there are viral sequences from 443 HIV-1 infected individuals associated with HIV-1 subtype C. The C subtype consensus sequence (C_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BE1.ID#:** These 6 sequences are from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. Accession numbers AJ228175, AJ228185, AJ228193, AJ228198, AJ228208 and AJ228228.
- 2) **BE2.VI358:** This subtype C sequence from Belgium was sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. Accession number X96531.
- 3) **BI1.ID#:** These eight sequences are from Burundi. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor-joining tree of the V3 region and is presented as an undetermined subtype in the section of the compendium [Penny (1996)]. Accession numbers U39252 and U39233, 91BU001; U39248 and U39237, 91BU002; U39239 and U39242, 91BU003; U39241 and U39243, 91BU004; U39240 and U39257, 91BU005; U39244 and U39246, 91BU006; U39245, U39247 and U39249, 91BU007; U39250 and U39251, 91BU008; U39253 and U39254, 91BU009 is possibly subtype CD recombinant.
- 4) **BR.19c:** This subtype C sequence is from Brazil. This study describes a case of heterosexual and subsequent mother to infant transmission of 2 HIV-1 subtypes, B and C [Janini (1998)]. DNA sequence analysis of pol, gag and env genes confirmed the presence of subtypes B and C in 3 family members. Accession numbers for env, gag and pol genes of both subtypes are U83689–U83699. Subtype C env accessions are U83690, U83692 and U83694.
- 5) **BR.W2BR025:** This sequence is part of a gp160 sequence from an asymptomatic individual from Brazil, sampled in 1992. A clone was derived from an expanded viral culture, expressed and sequenced. This sequence was provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences from this patient can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf (1994)]; [Osmanov (1994)]; [for HIV Isolation (1994)]. Accession number U15121. This sequence is from a 23 year old male hemophilia patient from Porto Alegre, Brazil. He had seroconverted 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 HIV isolate exhibited an NSI phenotype, when assayed by the WHO [Gao (1996a)]. 92BR025 is similar to another WHO sample; 91BR015. Accession number U52953. Entries with accession numbers U08720, U08785, U09126, U09132 and U09133 are also from W2BR025. An entry from India with accession number AF148246 is 97% identical to BR025 and most likely is a PCR contaminant, rather than an Indian sequence.
- 6) **BR.HSP203** Although this sequence is listed as unpublished in the database, it seems to be an extension of work published in [Morgado (1994), Sabino (1994c), Sabino (1996)]. It is from Sao Paulo, Brazil. Accession number U31585.
- 7) **BR2.91BR015:** This sequence is from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. It is from an adult AIDS patient. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two clones were sequenced. 91BR015 is similar to another WHO sample; 92BR025 [Penny (1996)]. Accession numbers U39234 and U39238.
- 8) **BY1.ID#:** These 3 sequences are from Byelorussia [Lukashov (1995)]. Accession numbers: L38410, BLR9A; L38409, BLR8A; L38408, BLR5A.
- 9) **BW1.ID#:** These 8 sequences are from a set of 23 complete genomes, several from each of the 8 patients, from Botswana [Novitsky (1999)]. 96BW17 is divergent in the pol gene, but still within subtype C (although an outlier). The pol gene of 96BW17 is similar to the pol gene from two isolates from Zimbabwe (Z1226 AF083262, and Z84 AF083267). The other 7 sequences are all clearly subtype C throughout the genomes. Accession numbers AF110959–AF110981.
- 10) **CD1.ID#:** This sequence is subtype C from the Democratic Republic of Congo (formerly Zaire) (Reitz, M. et al, unpublished 1995). Accession number U43102.

- 11) **CM11.ID#:** These 2 sequences are from 1994-1995 samples from 211 Cameroonian AIDS patients [Takehisa (1998)]. Of the 43 HIV isolates sequenced, 17 were subtype A, 1 was subtype B, 2 were subtype C and 1 was subtype G. Accession numbers AF023082, AF023084.
- 12) **CN1.CHI25:** This sequence is from the Guangxi Province, China from a blood sample collected in 1996 [Chen (1999)]. Subtypes B, D and CRF01(AE) were also identified in this study. Accession number AF080192.
- 13) **CU.93CU051:** This sequence is from Cuba (Gomez, C.E.G.R. unpublished 1997). Accession number Y14414.
- 14) **CY.HO021:** This is a sequence from a 51 year old woman whose husband had died of AIDS. She was born and lived in Zambia, before moving to Cyprus. She was asymptomatic, with a CD4 count of 200, and she had been seropositive for at least 6 years. This sample, like others in this study (see also subtypes A, B, F and I) was collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from patient 02 were sequenced [Kostrikis (1995)]. Accession numbers U28321 and U28661.
- 15) **DJ1.ID#:** These two sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced [Louwagie (1995)]. Accession numbers L22940 and L23065.
- 16) **ET2.ID#:** These nine sequences of Ethiopian isolates were part of a cohort considered to be heterosexually infected. PBMC DNA was PCR amplified and directly sequenced [Salminen (1996)]. For patient 2220, who had slim disease and AIDS, almost full length HIV-1 genome was PCR amplified from PBMC DNA and cloned. Proviral DNA from others in the cohort was PCR amplified and directly sequenced. Many other sequences from this set were less than 170 bp in length and not included here. All patients were from Addis Ababa and has slim disease and were classified as having AIDS by the WHO definition. Other regions of the genome are available for these isolates as well. Accession numbers U45481–U45502 (V3 region, all but 8 were too short to use), U15060–U15066 and U45503–U45504 (LTR NF-kB/NRE regions), M64001–M64009 (gag p7 region), M64015–M64018 (env gp41 region) and U46016 (C2220 complete genome).
- 17) **ET3.ID#:** These 94 sequences are from a study that describes the distribution of HIV-1 subtypes in Ethiopia. HIV-1 RNA was collected from sera (mostly from asymptomatic individuals) and a 284bp fragment covering the V3 region was amplified by RT-PCR and directly sequenced. All sequences were subtype C except for one subtype A [Abebe (1997)]. Accession numbers U88727–U88755, U88757–U88821.
- 18) **ET4.ID#:** These 3 sequences are from a study that describes the distribution of HIV-1 subtypes in Ethiopia, [Sherefa (1997)]. Several other sequences which did not cover the full V3 region were also part of this set. A few other sequences by Sherefa et al were listed only as being from Africa and have not yet been included in this analysis. Accession numbers U56367, U56368 and U56369.
- 19) **FR.MANJP:** This sequence is from France (Roques et al, unpublished 1998). Accession number AJ006746.
- 20) **FR1.ID#:** These 14 sequences are from members of the French military who are believed to have been infected while deployed outside of France (in Djibouti). An additional sequence (FRMP040, U58787) was listed as subtype C in the paper, but clustered with subtype B in analysis done at the HIV Database and it has been listed as subtype unclassified until more information is available. Other sequences from this study were subtypes A, B, E, and F [Lasky (1997)]. Accession numbers for subtype C were U58785, U58786, U58788–U58799.
- 21) **FR2.ID#:** These 2 subtype C sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers for subtype C are Z95462 and Z95464.
- 22) **GA.G134:** This sequence is from Gabon. G134 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon. Method of proviral DNA isolation was not described. DNA was PCR

- amplified and cloned. One clone per isolate was sequenced [Delaporte (1996), McCutchan (1996b)]. Accession number X90912. See also subtypes A, D, F, G, O and untyped/recombinant sequences from this same study.
- 23) **GB.00513**: This sequence is from the British isolate 93–00513. [Arnold (1995c)]. Accession number U21099.
 - 24) **GM.GM3**: This sequence is from Gambia (Bobkov et al, unpublished 1996). Accession number U33098. See also Gambian sequences of subtypes B and J.
 - 25) **GR1.ID#**: These 2 sequences are from Greece [Papa (1998)]. Subtypes A, B and D were also noted in this study. Accession numbers AF049300 and AF049301.
 - 26) **IL1.ID#**: These 6 subtype C sequences are from Israel [Gehring (1997)]. Other subtypes found in Israel in this study were A, B, D, K and CRF02(AG). Accession numbers for the subtype C sequences are X94387–X94390, X94393 and X94394.
 - 27) **IN1.ID#**: These four sequences are from samples from high risk patients in India, PCR clones, DNA, PBMC culture [Dietrich (1993)]. Accession numbers L07651 and L07653–L07655, X65638–X65640 and X68406.
 - 28) **IN2.D-ID#**: These five sequences are from samples from high risk patients in India, primarily WHO stage I. They were nested PCR amplified from DNA obtained from uncultured PBMC from patients serologically defined as HIV-1/HIV-2 mixed infections [Grez (1994)]. Accession numbers U07098 and U07100–U07103.
 - 29) **IN3.ID#**: These 8 sequences were isolated from Pune and New Delhi, India. All 8 sequences were from heterosexually infected patients from New Delhi, or Pune, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced [Tripathy (1996)]. Accession numbers U29179, U29694–U29698, U31362 and U31363. See also B_IN.IND9.
 - 30) **IN4.ID#**: These 24 sequences are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive STD patients (1, 3, 5, 6, 7, 11, 12, 13, 19, 20, 23, 29, 33), spouses of infected men (2, 4, 10, 16, 22, 26, 27), female prostitutes (36, 37) or bisexual men (8, 32). Two homosexual men from the same region had subtype A HIV-1 (see A_IN1.9 and 14) [Cassol (1996)]. Accession numbers U53278–U53285, U53287–U53290, U53292–U53303.
 - 31) **IN5.ID#**: These 7 sequences are from India. Another sequence from this publication was subtype A (see A_IN2) [Tsuchie (1993)]. Accession numbers D13420–D13424, D13426, D13427.
 - 32) **IN6.ID#**: These 7 sequences are from India. Another sequence from this publication was subtype A (see A_IN3) and two were of the B-prime Asian subclade of subtype B [Maitra (1999)]. Two of the subtype C samples were from child-mother pairs and the child sequence is presented here (AF101119 child, AF101118 mother; AF101126 child, AF101125 mother). Accession numbers AF101114, AF101115, AF101117, AF101119, AF101120, AF101123 and AF101126.
 - 33) **IN7.ID#**: These 24 sequences are from India. Six other subtype C sequences in this set were not included here because they were very similar to one or more of these that were included, the accession numbers of those not included follow the similar ones that were included in parentheses (Seth et al, unpublished 1999). Other sequences in this set were subtypes A and B (see A_IN4 and B_IN3). The 47016 sequence with accession number AF148246, is 97C sequence BR025 (U09133) and is most likely a PCR contaminant with this lab strain, rather than an authentic Indian sequence. Accession numbers AF148261 (AF148229), AF148239 (AF148238), AF148248, AF148260, AF148228 (AF148244), AF148252, AF148234, AF148254, AF148255, AF148233 (AF148232), AF148256, AF148251, AF148231, AF148242, AF148230, AF148259, AF148247, AF148235, AF148258, AF148253, AF148243, AF148237 (AF148236), AF148241 (AF148240), AF148257.
 - 34) **IN8.ID#**: These 5 sequences are from complete genomes from India [Lole (1999)]. One of the genomes was found to be recombinant between subtypes A and C. Accession numbers AF067154, AF067155, AF067157 and AF067158.
 - 35) **KE.NA113**: This sequence was derived from a patient who was part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The

- PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and D. [Zachar (1996a)]. Accession number U33762.
- 36) **KE1.ID#:** These 4 subtype C sequences are from Kenya [Robbins (1999)]. The samples were collected in 1994 and 1995 from male truck drivers and female sex workers near Mombasa and Nairobi. Subtypes A and D were also found in this study. Accession numbers for the subtype C sequences in this study: AF103919, AF103925, AF103932, AF103933.
 - 37) **KR1.ID#:** These sequences are from South Korea. Subtypes A, B and H were also reported, but the subtype H sequence (KR68), one of the subtype A sequences (KR61) and a subtype C sequence (KR75) were not submitted to the databases [Kim (1999)]. Accession numbers for the entire set are Z92548–Z92668. Subtype C accession numbers are Z92551, Z92552.
 - 38) **KE1.ID#:** These 7 sequences were derived from patients who were part of a study of breastfeeding women from Nairobi, Kenya. Viral DNA was amplified from uncultured patient PBMC, and the envelope V1-V5 region was sequenced after cloning into M13 phage. Other patients from this study had viral subtypes A, D, G and various recombinant forms [Neilson (1999)]. Accession numbers AF101458 and AF101460–AF101465.
 - 39) **LB.LE15:** This subtype C sequence is from a study of 25 sequences from a study of multiple HIV-1 subtypes in Lebanon [Pieniasek (1998)]. Of the 26 samples, 25 were classified as HIV-1: 10 were subtype B, 10 were subtype A, 1 was subtype C, 1 was subtype D and 3 were recombinant or untyped. The other sample was classified as HIV-2 subtype B. Accession number AF025697 is HIV-1 subtype C.
 - 40) **MW.SH750:** This sequence is from cloned PCR amplified cocultured PBMC DNA. SH750 was a black male sampled in the gold mines in Malawi in 1989. [Becker (1995)]. Accession number U06719.
 - 41) **MW1.ID#:** These 13 sequences are from pregnant women with risk factors from Malawi. PCR-direct, peripheral blood DNA [Orloff (1993)]. Accession numbers L07427, L07428, L07430–L07441 and L15721–L15735.
 - 42) **MW2.ID#:** These three sequences are from individuals from Malawi, generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. The C2V3 was excised from full gp160 sequences, derived from clones from expanded culture stocks. The sequence ID numbers are abbreviated, for example D3MA959 can be read as DAIDS sequence (D), isolated in 1993 (3), Malawi (MA), patient 301959 (959). Accession numbers U08453–U08455.
 - 43) **MY1.ID#:** These 2 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. Both of these prisoners had received medical treatment in India which included blood transfusion and organ transplants, and it is likely that they were infected in India. [Brown (1996)]. This report also included subtypes B and CRF01(AE) in Malaysia. Accession numbers U65549–U65550.
 - 44) **MZ1.ID#:** These 7 subtype C sequences are from a study of seven heterosexual patients residing in Maputo, Mozambique [Engelbrecht (1998)]. Blood was obtained in June 1996. Envelope V3 region was directly PCR amplified from uncultured PBMCs. The 300 bp PCR product was directly sequenced. Accession numbers AF045628–AF045634.
 - 45) **NL1.ID#:** These 2 sequences are from a Dutch woman whose partner was a recent immigrant to The Netherlands from Democratic Republic of Congo (formerly Zaire), and from a recent immigrant to The Netherlands from either Zambia or Democratic Republic of Congo (formerly Zaire). The first two letters of the ID# represent the two letter country code for the previous residence of the patient (UN = unknown). The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. The sequence from the Dutch/Democratic Republic of Congo (formerly Zaire) patient is not unequivocally a subtype C sequence, it may be subtype A or unclassifiable [Lukashov (1996)]. Accession numbers L76908, NL9402418; L76898, UN9305091.
 - 46) **NL2.ID#:** These 6 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions.

- 47) **NO1.ID#:** These 4 sequences are from Norwegian patients who were part of the Oslo HIV cohort study [Engelstad(1996)]. Uncultured PBMC DNA was PCR amplified in two nested PCR reaction steps. PCR products were directly sequenced. Where two peaks of equal height were observed at a single position, IUPAC ambiguity codes were used. Health, sex, year of sample (1989-1992), and risk group (IVDU, Het, Homo, Hemo) for each patient were noted in a table in the publication. The four subtype C sequences were all from heterosexuals who had sex with persons from Zambia (3) or South Africa (1). Thirty-six subtype B sequences were also part of this set (see B_NO1.ID#). Accession numbers X92912, X92913, X92916 and X92917.
- 48) **NZ1.ID#:** These 2 subtype C sequences are from New Zealand. Out the ten strains sequenced, 8 were subtype B and 2 were subtype C [Dwyer (1998)]. Accession numbers AF052630, AF052631. Both NZ3 and NZ7 were females infected heterosexually by partners who had previously lived in Africa.
- 49) **RU.ID#:** These 4 sequences are from Russia [Lukashov (1995)]. Accession numbers: L38418, RUS20A; L38404, RUS2A; L38406, RUS1A; L38414, RUS13A.
- 50) **RU1.YAN4:** This sequence is from Russia (Bobkov, V. et al, unpublished 1996). Accession number U33109.
- 51) **RW1.ID#:** These two sequences are from many clones from mothers infected in Rwanda. Mulder-Kampinga et al., Unpublished (1996) and [Kampinga (1997)]. Mother 566 was apparently dually infected with HIV-1 subtypes A and C. See also A_RW4.ID#. Accession numbers for mother 566 are Z76183, Z76188–Z76197, Z76218–Z76231, Z76284–Z76293, Z76295–Z76299, Z76301, Z76303–Z76311, Z76322–Z76342, Z76725–Z76732 and Z76734; for mother 134, Z76039–Z76042, Z76044.
- 52) **SE1.ID#:** These 16 sequences are from a study of 270 HIV-1 infected individuals living in Sweden [Alaeus (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 bloods samples were available from 33 men and 42 women, from 15 different African countries. The 16 subtype C sequences were from individuals who were thought to have been infected in Ethiopia (SE9279 U76174, SE8899 U76161 and SE7854 U76122), Eritrea (SE7410 U76128 and SE8848 U76166), Somalia (SE8879 U76169), Tanzania (SE8684 U76157, SE9085 U76173 and SE7564 U76184), Uganda (SE7159 U76117), Botswana (SE9283 U76176), Zambia (SE8056 U76120), Zimbabwe (SE8565 U76148, SE8337 U76114, SE9337 U76178 and SE9338 U76179), Kenya (SE8890 U76171) and Mozambique (SE6077 U76123). SE8337 and SE9337 are not included in this alignment because they are from the sex partners of SE8565 and SE9338 respectively, and are very similar to those sequences.
- 53) **SE2.ID#:** These 2 subtype C sequences are from a study that includes the analysis of HIV-1 strains in seven cases of mother-to-child transmission in Sweden [Contag (1997)]. HIV-1 isolation from the mothers was conducted during the three trimesters, a few days after delivery and 6 months after parturition. The children were born between June 1989 and August 1992 and were followed by virus isolation at birth and at 3, 6, 12 and 18 months of age unless circumstances called for an extra sample. Breast-feeding was denied by all of the women described. See also A_SE.H4, B_SE5.ID#, D_SE.H3 and AE_SE.H1. Accession Numbers U56263-U56335.
- 54) **SE3.ID#:** These 3 subtype C sequences are from Sweden. Several clones were sequenced for each patient, sampled over time since primary infection took place [Karlsson (1999)]. The authors compared the diversity of sequences in primary infection with samples from subtypes B, C and CRF01(AE). Accession numbers for the entire set are AF014075–AF014104 and AF068484–AF068533.
- 55) **SG1.ID#:** These 3 sequences are from Singapore. Other sequences in this set were subtypes A, B and CRF01(AE) [Se-Thoe (1998)]. Accession numbers AF004234, AF004247 and AF004257.
- 56) **SN.SE364:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession number L22944.
- 57) **SN.P1581:** This sequence is from a study done on individuals infected with non-B clade virus who were randomly obtained from a cohort of registered sex workers in Senegal, West Africa. PBMC were separated, cryopreserved and shipped to USA for CTL studies [Cao (1997)]. Of the 14 sequences evaluated 10 were subtype A, three were subtype G and 1 was subtype C. Accession number AF020822.

- 58) **SN1.ID#:** These 3 subtype C sequences are from Senegal [Kanki (1999)]. Subtypes A, D, G, CRF02(AG) and another AG recombinant were also found in Senegal. Accession numbers for the subtype C sequences are AF085304, AF085315 and AF085320.
- 59) **SO.1574:** This sequence is from a study of blood and CSF samples taken from the Somali patient 1574, CDC classification II [Keys (1993)]. Accession numbers Z23188, Z23190–Z23191, and Z23228–Z23231.
- 60) **SO.SM145:** A Somali sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession number L22946.
- 61) **TW.252** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, E, F or G. [Chang (1997)]. Accession number U73055.
- 62) **TZ1.ID#:** This sequence is from a set of 15 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing [Zwart (1993)]. The other 13 samples were subtypes A (4) and D (10). Accession numbers L01301–L01302.
- 63) **TZ2.ID#:** These 47 sequences are part of a set of 86 sequences from samples collected from symptomatic AIDS patients in December 1995 at Mbeya Referral Hospital in southwest Tanzania. Uncultured PBMC DNA was PCR amplified and directly sequenced. Serotyping was also done on all samples to test the ability of serology to subtype these A, C, D and recombinant HIV-1 isolates [Hoelscher (1997), Hoelscher (1998)]. The sequences have not yet entered the databases (12-23-99).
- 64) **TZ3.ID#:** These 31 sequences are from Dar es Salaam on the eastern coast of Tanzania [Renjifo (1999)]. Subtypes A, D and many independently-derived recombinants were also found in this study. Accession numbers for the entire set are AF038051–AF038121 and AF106332–AF106472. Accession numbers for sequences which are subtype C in the V3 region are AF038076–AF038093, AF038095–AF038097, AF038118 and AF106350–AF106358.
- 65) **UG.45:** A single sequence used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda [Pestano (1995)]. Accession number U11597. See also A_UG1.964, B_US17.ID#, and D_UG7.ID#.
- 66) **UG.UG268:** A Ugandan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession number L22948.
- 67) **UG1.ID#:** These 3 sequences are from Uganda. Blood specimens from 1100 patients were collected in 5 districts of Uganda, out of these 739 were selected for further subtyping in env or pol regions. Subtype A and D specific probes were used to type the C2V3 region. Subsequent sequence analysis of 19 randomly selected specimens revealed subtypes D(n=16), C(n=3). Accession numbers AF016329, AF016330 and AF016331.
- 68) **ZA.NOF:** This sequence is from a South African individual who was part of a study of HIV-1 strains in India. This sequence was found to be closer to the Indian sequences, than are other isolates from Africa. PCR amplified DNA from PBMC cultures were sequenced [Dietrich (1993)] and [Becker (1995)]. Accession numbers L07426 and U06716. See also C_IN3.ID#, B_IN.IND9.
- 69) **ZA1.ID#:** These 2 sequences are from 2 individuals in South Africa. ZA514 was from a 59 year old mixed-race male heterosexual with AIDS and the virus was NSI. ZA517 was from a 33 year old mixed-race male heterosexual with ARC and the virus was NSI. The samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product [Engelbrecht (1995)]. Accession numbers L48067–L48068. Database entries U33781 and U33782 are shorter env gene fragments from these same clones. See also B_ZA and D_ZA sequences from this same study.
- 70) **ZA2.ID:** These 3 sequences are from clones from PCR amplified cocultured PBMC DNA. Dlu was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. Gom was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. BooyD was a mixed-race mother sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. A short sequence from a

seropositive child of BooyD is presented in entry U07015 [Becker (1995)]. Accession numbers U07237, U06717 and U06718.

- 71) **ZA3.ID#:** These 44 subtype C sequences are from a study of gold miners from Westonaria, a district situated 50km from Johannesburg (South Africa) [Bredell (1998)]. The men were 18-65 years of age and employed as migrant workers living in single-sex hostels on the mines. Serostatus at the time of employment, date of seroconversion and geographical place of infection are unknown. Samples 95ZA853BWA and 96ZA119BWA were from same patient taken 2 months apart. 19 of 43 individuals were from South Africa, 13 were from Lesotho, 5 were from Botswana and 3 each were from Swaziland and Mozambique. Accession numbers AF053181–AF053224.
- 72) **ZA4.ID#:** These 21 subtype C sequences are from a study of 72 HIV-1 seropositive women from the KwaZulu-Natal region of South Africa [Moodley (1998)]. The mean age of the women was 26 years. All but one of the women were asymptomatic with a mean CD4 count of 459 cells/ml. One patient (KZN149) had a CD4 count of 75 cells/ml and had AIDS. Data from this study shows the dramatic growth of HIV-1 subtype C in this population in South Africa. See also A_ZA.134 and B_ZA.0117. Accession numbers AF053277-AF053299.
- 73) **ZA5.ID#:** These 5 subtype C sequences are from unpublished mother-infant sequences by Loubser, A.S. and Williamson, C. and from sequences published in [Van Harmelen (1999)]. For the mother-infant pairs, only the infant sequence is reported here. Accession numbers AF095825, AF095826, AF095831 and AF095834–AF095838. Subtypes A, B and D were also found in this study in South Africa.
- 74) **ZM1.ID#:** Two sequences from Zambia, from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession numbers L22954 and L22956.
- 75) **ZW.2647T:** This sequence is one of several taken from blood and CSF samples taken from Zimbabwe patient 2647, CDC classification II [Keys (1993)]. Accession numbers Z23196–Z23199 and Z23236–Z23239.
- 76) **ZW1.ID#:** These 22 sequences are taken from blood samples from Zimbabwe recent seroconverters [Tien (1999)]. All were subtype C, and 3 of the 22 were syncytia-inducing (SI) in MT-2 cells. Accession numbers AF056117–AF056141.

D Subtype

At this time there are viral sequences from 182 HIV-1 infected individuals associated with HIV-1 subtype D. The D subtype consensus sequence (D_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BE2.VI358:** This subtype D sequence from Belgium was sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. Accession number X96525.
- 2) **BR.RJ100** This subtype D sequence was from a male of unreported age and unknown risk group from Rio de Janeiro, Brazil. He is thought to have been recently infected at the time of sampling for this sequence in September 1996, because he was a blood donor in early 1995. The patient's CD4 count was 96 cell/ml in September, 1996 and 290 cells/ml in January 1997, after antiretroviral therapy had been initiated [Morgado (1998)]. Accession number AF000238. In this study of 131 HIV-infected individuals, 106 were classified as subtype B by HMA and 20 were classified as subtype F. Only this one sample was subtype D, and it was the only one sequenced.
- 3) **CD.84ZR085:** This sequence is from a complete genome of a 1984 isolate from the Democratic Republic of Congo (formerly Zaire). The env region clusters with subtype D, and the pol gene is a subtype D outlier [for HIV Isolation (1994), Gao (1994a), De Wolf (1994)]. Accession number U88822.
- 4) **CD.ELI:** This sequence is from the Democratic Republic of Congo (formerly Zaire) isolate ELI [Alizon (1986)] and [Goodenow (1989)]. The complete genomic sequence and an infectious clone are available. In the 1995 Compendium (pages III-45 and III-47), ELI was listed as an unlikely D/A mosaic, with only gp41 being weakly A-like. ELI was cultured in 1983, from a 24 year old woman with AIDS. Accession numbers M27949, K03454 and X04414.
- 5) **CD.JY1:** This sequence is from Democratic Republic of Congo (formerly Zaire) isolate Z-84, clone JY1. [Yourno (1988)]. Accession number J03653.
- 6) **CD.MAD:** This sequence is from an asymptomatic Democratic Republic of Congo (formerly Zaire) woman who was seropositive for HIV-1 by several French-approved HIV tests and and by HIV-1 western blot. The viral sequences obtained for the env V3 region and gp41 show that this isolate belongs to M-group subtype D [Cohen (1995)]. Accession number X83216. A region of gp41 is also sequenced, see X83215.
- 7) **CD.NDK:** This sequence is from an infectious clone of the Democratic Republic of Congo (formerly Zaire) isolate NDK. The molecular clone is highly cytopathic in vitro [Spire (1989)]. The complete genomic sequence is available. Several lab-generated recombinants have been constructed and tested in order to locate the regions of this genome which contribute to its extreme cytopathogenicity. Accession number M27323.
- 8) **CD.Z2Z6:** This sequence is from an infectious clone from a Democratic Republic of Congo (formerly Zaire) isolate Z2. Theodore T, and Buckler-White A, unpublished. The complete genomic sequence is available. Accession number M22639. See also [Srinivasan (1987)]. The database entry with accession numbers K03458 and M16322, is from the same isolate, but not the same clone.
- 9) **CD1.ID#:** These four sequences are part of a set of 14 A and D sequences from women from Democratic Republic of Congo (formerly Zaire). 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. Sequences were determined by directly sequencing PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs [Potts (1993a)]. Accession numbers L19623, L19627, L19631 and L19635.
- 10) **CF.4020:** This sequence was the only D subtype from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. It is one of several cloned PCR products, derived from cultured proviral DNA [Murphy (1993)]. A full gp120 sequence from this isolate was kindly provided prior to publication by Dr. MP Kiény of Transgene, Strasbourg, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system [Richalet-Secordel (1994)]. The year of isolation and health status of individuals from which the viruses were isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kiény. Sequences of clones from sample 4093 (Accession numbers L11516, L11517) which was taken from the

same patient as 4020 one year later, are closely related and are not included here. Accession numbers L11472, L11473, U43138.

- 11) **CF.T22:** This sequence is from a set of sequences obtained from patients from the Central African Republic [Muller-Trutwin (1999)]. Accession number AF067760.
- 12) **CG.POA:** This sequence is from a 1988 sample collected in Brazzaville, Congo [Candotti (1991), Candotti (1999)]. The patient was an adult with AIDS at the time of sampling. Subtypes A, D, F, G and recombinants (AG and AE) were identified in this study. Accession number for the subtype D sample is AF082318. Gag sequences are also available for some of the samples, with accession numbers M73472–M73480.
- 13) **CI.CI-13:** A single D subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-13 was symptomatic, and serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 2 clones were sequenced, and the X72028 clone is presented here. [Janssens (1994a)]. Accession numbers X72028–X72029.
- 14) **CII.ID#:** These 2 sequences are from Abidjan, Ivory Coast. No other information is yet available (Ellenberger D.L. unpublished 1997). Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000469, AF000451.
- 15) **CM.CMR61D:** CMR61 is one of two D group sequences from a 23 year old female sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa (1997b)]. The other sequence from the same patient is labelled as CMR709 and is from a separate blood sample. Accession numbers U58149, U58151 and AF097696.
- 16) **CM2.ID#:** These four sequences are all from Cameroon [Takehisa (1998)]. Accession numbers U70013, U70012, U70011, U70009.
- 17) **CM3.ID#:** These 2 subtype D sequences are from samples collected in 1995 from 53 northern Cameroonian Muslim AIDS patients [Mboudjeka (1999)]. Of the 16 HIV patients sequenced, 5 were subtype A, 6 were CRF02(AG) IbNG-like, 2 were subtype D, 1 was subtype G and 2 were subtype H. Accession numbers for subtype D are AF028318 and AF028329.
- 18) **CN1.CHI09:** This sequence is from the Guangxi Province, China from a blood sample collected in 1996 [Chen (1999)]. Subtypes B, C and CRF01(AE) were also identified in this study. Accession number AF080207.
- 19) **FR1.ID#:** These 9 subtype D sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers Z95450, Z95466, Z95467, Z95468, Z95469, Z95470, Z95471, Z95472 and Z95473.
- 20) **GA1.ID#:** These 2 sequences are from 1988 or 1989 samples from patients with AIDS living in Franceville Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced, [Delaporte (1996)]. Accession numbers X90919, G109; X90920 G141. See also subtypes A, C, F, G and O sequences from this same study. A gag gene from G109 is found with accession number L11766.
- 21) **GB.CPHL4:** This sequence is one of the British isolate 93–43424, clones 2, 6 and 35. It was referred to as 93–43424 in [Arnold (1995c)] and as CPHL4 in [Arnold (1995a)]. Accession numbers U21098 (clone 35) and U23121–U23122 (clones 2 and 6 respectively). CPHL4 is a female who is believed to have contracted the virus from CPHL5 through heterosexual contact, over 8 years prior to the date samples were collected for this analysis. Sequences from CPHL5 (U23123–U23125) are not included in this alignment due to this epidemiological relationship.
- 22) **GH.GH3:** This sequence is from Ghana. Subtypes A and an A/G recombinant were also detected in this study [Takehisa (1997a)]. Accession number U67049.
- 23) **GH1.ID#:** These 2 sequences are from Ghana. Subtypes A and G were also detected in this study [Ishikawa (1996)]. Accession numbers for subtype D are U67682 and U67684.
- 24) **GR1.ID#:** These 4 sequences are from Greece [Nasioulas (1998)]. Subtypes A, B and C were also noted in this study. Accession numbers AF049294, AF049295, AF049298 and AF049299.

- 25) **IL.PA:** This subtype D sequence is from Israel [Gehring (1997)]. Other subtypes found in Israel in this study were A, B, C, K and CRF02(AG). Accession number for the subtype D sequence is X94398.
- 26) **KE.NA116:** This sequence was derived from a patient who was part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and C. [Zachar (1996a)]. Accession number U33765.
- 27) **KE1.ID#:** These three patients were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced [Janssens (1994c)]. Accession numbers for subtype D sequences in this publication: U12984–U12986.
- 28) **KE2.ID#:** These 3 sequences are part of a set of 5 subtype D sequences from Kenya. One of the subtype D sequences was too short to include here (AF004886) and one (AF004887) is represented by a complete genome sequence (AF133821) in the KE3 set. Another 13 sequences from this set were found to be subtype A [Poss (1997)]. Accession numbers AF031660, AF004886, AF004887, AF004898, AF004899.
- 29) **KE3.ID#:** These 3 sequences were derived from patients who were part of a study of breastfeeding women from Nairobi, Kenya. Viral DNA was amplified from uncultured patient PBMC, and the envelope V1-V5 region was sequenced after cloning into M13 phage. Other patients from this study had viral subtypes A, C, G and various recombinant forms [Neilson (1999)]. Accession numbers AF101466, AF133821 and AF133822.
- 30) **KE4.ID#:** These 5 subtype D sequences are from Kenya [Robbins (1999)]. The samples were collected in 1994 and 1995 from male truck drivers and female sex workers near Mombasa and Nairobi. Subtypes A and C were also found in this study. Accession numbers for the subtype D sequences in this study: AF103909, AF103912, AF103916, AF103920, AF103927.
- 31) **LB.LE22:** This subtype D sequence is from a study of 25 sequences from a study of multiple HIV-1 subtypes in Lebanon [Pieniazek (1998)]. Of the 26 samples, 25 were classified as HIV-1: 10 were subtype B, 10 were subtype A, 1 was subtype C, 1 was subtype D and 3 were recombinant or untyped. The 26TH sample was classified as HIV-2 subtype B. Accession number AF025705 is HIV-1 subtype D.
- 32) **ML.103** This subtype D sequence is from a study that looks at the prevalence of different subtypes of HIV-1 and HIV-2 circulating in female sex workers in Bamako, the capital city of Mali [Peeters (1998)]. A total of 176 CSWs were tested and 81 were HIV infected. Of the 81, 63 were infected with HIV-1, 7 were infected with HIV-2 and 11 were dually infected with HIV-1 and HIV-2. HMA assays indicated that 80 percent of HIV-1 infections were with subtype A virus. Only 9 viruses, with ambiguous HMA results, were sequenced. Out of these 9 sequences one was subtype A, one was subtype D and 7 were subtype G. Accession number Y14362.
- 33) **NL1.ID#:** These 7 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. Accession numbers L76900, UG9307184; L76907, ZR9402261; L76892, ZR929193; L76904, UG9401525; L76895, AO9302187; L76872, ZR891183; L76876, ZR901100.
- 34) **RU.RUS14A** This sequence is from [Lukashov (1995)]. Accession number L38415.
- 35) **SE.H3B** This subtype D sequence is from a study that includes the analysis of HIV-1 strains in seven cases of mother-to-child transmission in Sweden [Contag (1997)]. HIV-1 isolation from the mothers was conducted during the three trimesters, a few days after delivery and 6 months after parturition. The children were born between June 1989 and August 1992 and were followed by virus isolation at birth and at 3, 6, 12 and 18 months of age unless circumstances called for an extra sample. Breast-feeding was denied by all of the women described. See also A_SE.H4, B_SE5.ID#, C_SE2.ID# and AE_SE.H1. Accession numbers U56263–U56335.
- 36) **SE1.ID#:** These 17 sequences are from a study of 270 HIV-1 infected individuals living in Sweden [Alaeus (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 blood samples were available from 33 men and 42 women, from 15 different African countries. The 19 subtype

D sequences were from individuals who were thought to have been infected in Democratic Republic of Congo (formerly Zaire) (SE6687 U76135 and SEu1352 U76116), Uganda (SE8681 U76156, SE8680 U76155, SE6274 U76130, SE6488 U76164, SE7565 U76124, SE7386 U76127, SE6184 U76139, SE6958 U76136, SE6339 U76129, SE8384 U76146, SE7076 U76134 and SE8603 U76153), Eritrea (SE8420 U76115 and SE8564 U76150), Kenya (SE9048 U76172 and SE9340 U76177) and Gambia (SE6095 U76163). SE6958 and SE8681 are not presented here because they are from the sex partners of SE6184 and SE8680 respectively. Accession numbers for the entire set of all subtypes are U76114-U76186 and L41176-L41179.

- 37) **SE.ID#:** These subtype D sequences are from Sweden [Leitner (1995)]. Direct sequencing of PCR products with ambiguity codes (R, Y, W etc) for heterogeneous sequence positions. Accession numbers L40746, L40747.
- 38) **SN.SE365:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession number L22945.
- 39) **SN1.ID#:** These 5 subtype D sequences are from Senegal [Kanki (1999)]. Subtypes A, C, G, CRF02(AG) and another AG recombinant were also found in Senegal. Accession numbers for the subtype D sequences are AF085285, AF085302, AF085312, AF085321 and AF085324.
- 40) **TZ1.ID#:** These ten sequences are from a set of 15 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing [Zwart (1993)]. Several clones from each sample were sequenced. The other 5 were subtypes A (4) and C (1). Accession numbers L01298–L01300, L01303–L01311, L01317–UL01334.
- 41) **TZ2.ID#:** These eight sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka (1994)]. Accession numbers U12406, U12407, U12410–U12415. Two other sequences in this set were subtype AD recombinant.
- 42) **TZ3.ID#:** These 2 sequences are from the Mara region of rural northwest Tanzania [Robbins (1996)]. Subtype A and C/D recombinants were also found in this study. Accession numbers U61880 and U65075.
- 43) **TZ4.ID#:** These 13 sequences are part of a set of 86 sequences from samples collected from symptomatic AIDS patients in December 1995 at Mbeya Referral Hospital in southwest Tanzania. Uncultured PBMC DNA was PCR amplified and directly sequenced. Serotyping was also done on all samples to test the ability of serology to subtype these A, C, D and recombinant HIV-1 isolates [Hoelscher (1997), Hoelscher (1998)]. The sequences have not yet entered into the databases (12-23-99).
- 44) **TZ5.ID#:** These 21 sequences are from Dar es Salaam on the eastern coast of Tanzania [Renjifo (1999)]. Subtypes A and C and many independently-derived recombinants were also found in this study. Sequences with accession numbers AF106359 and AF038099 are from the same patient, so only AF038099 is reported here. Accession numbers for the entire set are AF038051–AF038121 and AF106332–AF106472. Accession numbers for sequences that are either gag-D and env-D or just env-D (gag not sequenced) with no evidence of recombination are AF038099–AF038111, AF106359–AF106361, AF106363–AF106367 and AF106369–AF106372.
- 45) **UG.KIT3:** This sequence is from a Ugandan. One of several PCR-clones, peripheral blood DNA. Intact env sequences are available from this sample [Bruce (1993)]. Accession numbers M98408–M98416 and duplicate entries with accession numbers Z19524–Z19531, Z19533.
- 46) **UG.UG23:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was PCR amplified and cloned. This sequence is from an individual clone [Atkin (1993), Pestano (1993)]. Accession number M98504.
- 47) **UG1.ID#:** These twelve sequences are from asymptomatic individuals from Uganda, sampled in 1992. Each sequence is a one of several cloned PCR products derived from cell-cultured proviral DNA or culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 Compendium [De Wolf (1994)]; [Osmanov (1994)]; [Gao (1994b)]. Accession numbers U08721–U08741, U08786–U08787, U08803–U08809,

- U08821–U08824, U27399, U43386. A sequence not reported to be related to these Ugandan isolates is suspected to represent lab contamination of PCR reactions with one of these isolates: U58827 (FRMP153) is 99% identical to 92UG046.
- 48) **UG3.ID#:** These 10 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic. Blood samples were obtained in 1990. Each sequence is one of several cloned PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs [Albert (1992)]. Accession numbers L00614–L00618, L00733–L00737, M98894–M98899, M98901, M98906–M98907, M98911–M98913, M98918, M98920–M98923, M98929–M98937, M98942–M98945 and M98967–M98975. The M98911 sequence is not included here, because it is 99% identical to the UG274 sequence with accession number L22950.
 - 49) **UG4.ID#:** One to 4 clones of each these two Ugandan isolates were sequenced, but only one clone is shown here [Douglas (1996)]. Other Ugandan isolates sequenced in this study were subtype D/A recombinant. London subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. Accession numbers U36867, U36868, U36871, U36884–U36887.
 - 50) **UG5.ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individuals from which the virus was cultured was unspecified, and the year of viral isolation was 1990. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. Accession numbers L22947 and L22949–L22950. UG266 (L22947) was reported to be subtype A in gag.
 - 51) **UG6.ID#:** These three sequences are from AIDS patients living in Uganda. The year of viral isolation was 1987. Viruses were cultured with HUT-78 cells for an unspecified length of time. The V3 region of env (gp160) was amplified, cloned and sequenced. [vonBrunn (1995)]. Accession numbers U15005, U15006 and U15007.
 - 52) **UG7.ID#:** These sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. All patients were asymptomatic individuals reporting for regular blood drawing at the Nakasero blood transfusion service, Kampala, Uganda [Pestano (1995)]. Accession numbers U11595, U11596, and U11598. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.
 - 53) **UG8.#:** These 2 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending the clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4%) were all asymptomatic [Buonaguro (1995)]. Database accession numbers U44881 and U44884. Eight subtype A sequences were also found in this study (see A_UG5).
 - 54) **UG9.ID#:** These 16 sequences are from Uganda. Blood specimens from 1100 patients were collected in 5 districts of Uganda, out of these 739 were selected for further subtyping in env or pol regions. Subtype A and D specific probes were used to type the C2V3 region. Subsequent sequence analysis of randomly selected specimens revealed subtypes A (n=17), D(n=26) and C(n=3). Accession numbers AF016332–AF016347.
 - 55) **US.AMK:** This sequence comes from a student living in Alabama, who moved from Democratic Republic of Congo (formerly Zaire) to the US in 1988. Virus was isolated from the patient's PBMCs in 1993; this isolate was PCR amplified, and amplification products from both gag (U08192) and env (U27419) were subcloned and sequenced. His CD4 count was < 5 cells/mm³, and he was symptomatic at the time of viral isolation [Gao (1994b)] and [Gao (1996a)]. AMK is also known as 93ZR001. Accession numbers U08193, U27419.
 - 56) **ZA.034:** This subtype D sequence is from South African sequences published in [Van Harmelen (1999)]. This sequence seems to be a subtype D outlier, and may be from a recombinant virus. Accession number AF095827. Subtypes A, B and C were also found in this study in South Africa.
 - 57) **ZA1.ID#:** These 5 sequences are from 5 individuals in South Africa. ZA500 was from a 41 year old white male homosexual with ARC and the virus was non-syncytium-inducing (NSI). ZA501 was from a 24 year old white male bisexual with AIDS and the virus was SI. ZA505 was from a 36 year old white male homosexual with AIDS and the virus was SI. ZA506 was from a 33 year old white male homosexual with AIDS and the virus was SI. ZA507 was from a 37 year old white male homosexual with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene

was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product [Engelbrecht (1995)]. Accession numbers L47608, L48061, L48062, L48070 and L48072. Database entries U33769, U33771–U33773 and U33780 are shorter env gene fragments from these same clones. See also B_ZA and C_ZA1 sequences from this same publication.

F Subtype

At this time there are viral sequences from 84 HIV-1 infected individuals associated with HIV-1 subtype F. The F subtype consensus sequence (F_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published.

- 1) **AR1.ID#:** These two sequences are from direct sequencing of PCR products from uncultured PBMCs, from 1993 samples from Buenos Aires, Argentina. Patient 21280 had AIDS and reported IV drug abuse. Patient 20016 was asymptomatic and HIV risk behavior was unknown. Two other samples taken from unrelated patients in 1993 were subtypes B or B/F recombinant [Marquina (1996)]. Accession numbers U68522 and U68524.
- 2) **AR2.ID#:** These 3 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were also tested by HMA [Campodonico (1996)]. Accession numbers U37032, U37033 and U37043.
- 3) **AR3.ID#:** These 12 sequences are from Argentina. A total of 25 patients were included in this study. Of the 25 sequences determined, 12 were found to belong to subtype F1 and 12 were found to belong to subtype B, and one was B/F1 recombinant [Fernandez-Medina (1999)]. Accession numbers AF155513, AF155517, AF155521–AF155523, AF155527–AF155530, AF155532, AF155533 and AF155535.
- 4) **BE1.ID#:** These 5 sequences are from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. Accession numbers AJ228148, AJ228210, AJ228222, AJ228225 and AJ228230.
- 5) **BE2.VI850:** This subtype F1 sequence from Belgium was sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. The sequence is from a Belgian man whose wife was infected in the Democratic Republic of Congo. Accession number X96527. A complete genome is available with accession number AF077336.
- 6) **BO1.ID#:** These two sequences are from Bolivia (Velarde-Dunois et al, unpublished 1998). Accession numbers AF031950 and AF031952.
- 7) **BR.7944:** This sequence represents a single env F subtype sequence found among 22 Brazilian outpatients with varying degrees of disease progression. It was identified by Potts et al. as the single sequence which did not cluster with North American sequences in phylogenetic analysis. Consensus, PCR clones, peripheral blood PBMC DNA. [Potts (1993b)]. Accession number L19237.
- 8) **BR.RJI03:** An F subtype sequence from Rio de Janeiro, Brazil. 26 additional B and a B-F recombinant were also observed in this set. Year of isolation for RJI03 was 1993, from a woman of CDC clinical stage II. [Morgado (1994)]. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II disease, and the PCR product was directly sequenced. Accession number U00422. See also [Sabino (1994c)] Accession number U08974.
- 9) **BR1.ID#:** Three sequences from Brazil of the F1 subtype. Full length env (gp160) was amplified from proviral DNA of cultured PBMCs, cloned and sequenced [Louwagie (1994)]. Accession numbers L22082, L22084 and L22085. The gag gene of these same isolates is found in L22083, L22086 and L11751. Although shown in figure 2B of the paper as clustering with subtype F in the gag region, BZ126 (L22083) seems to be a subtype A outlier with a strong similarity to subtype C at the 3' end, and is 99% identical to the subtype A/C recombinant isolate ZAM184. See also B_BR4.BZ-ID#.
- 10) **BR2.ID#:** These 5 sequences are from Sao Paulo Brazil. They seem to be an extension of work published in [Morgado (1994)] and [Sabino (1994c)]. They are proviral DNA sequences from cloned PCR products, from uncultured PBMCs [Sabino (1996)]. Accession numbers U31588, U31592–U31595.
- 11) **BR3.93BR02017:** This sequence is part of a set of sequences generated through the WHO Global Programme on AIDS. The virus was derived from a 52 year old asymptomatic male, from Rio de Janeiro, Brazil, whose route of infection is thought to be due to bisexual contact. The blood sample was taken in 1993. The env sequence clusters with HIV-1 subtype F1 sequences. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao (1996a)]. Accession number U27401. A complete genome of the same isolate is available with accession number AF005494.

- 12) **BR4.ID#:** These 6 sequences are from study of asymptomatic blood donors in Rio de Janeiro Brazil [Tanuri (1999)]. One of the 42 sequences reported was subtype B in env and F1 in gag. Two were subtype F1 in env and B in gag. Thirty-two were subtype B in both env and gag. One was subtype D in env V3 and gag, but subtype B in the env gp41 region. Accession numbers for the six subtype F1 are AF034003, AF034006–AF034009 and AF034031.
- 13) **BR5.ID#:** These 4 sequences are from Brazil (Bongertz, V. et al, unpublished 1998). Accession numbers AF062422–AF062425.
- 14) **BR6.ID#:** These 5 sequences are from Brazil (Vincente, A.C.P. et al, unpublished 1999). Accession numbers AF076314, AF076315, AF076317, AF076320, AF076321.
- 15) **BR7.ID#:** These 4 sequences are from Brazil, [Ramos (1999)]. Accession numbers AF113566–AF113568 and AF113575.
- 16) **BR8.ID#:** These 2 sequences are from Brazil (Couto-Fernandez, J.C., unpublished 1999). Accession numbers Y18756 and Y18758.
- 17) **CG.C12:** This sequence is from a 1992 sample collected in Brazzaville, Congo [Candotti (1991), Candotti (1999)]. The patient was an adult with AIDS at the time of sampling. Subtypes A, D, F, G and recombinants (AG and AE) were identified in this study. Accession number for the subtype F sample is AF082305. Gag sequences are also available for some of the samples, with accession numbers M73472–M73480.
- 18) **CM1.CA-ID#:** These sequences are 3 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals; specifically patients CA16 and CA20 were asymptomatic and patient CA4 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. The F subtype designation of these sequences is tentative. Although the F subtype sequences from Cameroon and Brazil consistently form a clade in phylogenetic analyses, the branch lengths between isolates from the two countries are typical of inter-subtype distances, and sequences from the two countries each form their own distinct clade within the F subtype (HIV database and Wouter Janssens, personal communication) [Nkengasong (1994)]. Accession numbers for the subtype F sequences are X80443, X80448 and X80451.
- 19) **CM2.ID#:** These three sequences are all from Cameroon [Takehisa (1998)]. Accession numbers U69997–U69999.
- 20) **CM3.ID#:** These 3 sequences are from Cameroon [Triques (1999)]. Accession numbers AJ237801, AJ237805 and AJ237806. Gag genes are also available with accession numbers AJ237793–AJ337795.
- 21) **FL.9363:** This subtype F sequence is from Finland (Laukkanen, T., unpublished 1999). Accession number AF075703.
- 22) **FR.BCB85:** This subtype F sequence is from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession number Z95465.
- 23) **FR1.ID#:** This sequence is from a member of the French military who is believed to have been infected while deployed in the Central African Republic in 1992. Other sequences from this study were subtypes A, B, C, E and unclassified [Lasky (1997)]. Accession number U58807.
- 24) **FR3.MP411:** This sequence is from France [Triques (1999)]. Accession numbers AJ237804, AJ249238. Gag genes are also available.
- 25) **MQ1.ID#:** These 4 sequences are from Martinique [Desgranges (1996)]. Subtype B was also found. Accession numbers for subtype F are U67707–U67710.
- 26) **NL1.ID#:** These 2 sequences are from recent immigrants to The Netherlands from Brazil and the Democratic Republic of Congo (formerly Zaire). The first letter of the ID# represents the country for the previous residence of the patient. The first two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. The patient referred to a 890819 by Lukashov is apparently the same as patient A12 in [Wolfs (1992a)], because the sequences are 99Patient A12 was the donor in a donor-recipient pair studied by Wolfs [Wolfs (1992a)]. The recipient accession numbers are M91849–M91856. Accessions M91840–M91847 seem to be B/F recombinants.
- 27) **NL2.ID#:** These 2 sequences are from The Netherlands [Lukashov (1996)]. Accession numbers L76871, ZR890819; L76901, BR9400960

- 28) **NL2.ID#:** These 4 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions. Accession numbers for the subtype F sequences are F032172, AF032181 and AF032212.
- 29) **RO.BC11:** This sequence is from Romania [Triques (1999)]. It seems to be different from the RO2.BC11 patient in the RO2 set below. Accession number AJ237792.
- 30) **RO1.ID#:** These nine sequences are from isolates from Romanian children, in different clinical stages. All isolates showed cytopathic properties in peripheral blood mononuclear cells. DNA sequences are direct sequences of PCR products amplified from co-cultured PBMCs. The patients are also known as RM(A-J) [Dumitrescu (1994)]. Accession numbers L19571–L19579.
- 31) **RO2.ID#:** These fifteen sequences are from isolates from Romanian children, infected nosocomially, and adults, infected through heterosexual transmission or transfusion. DNA sequences are direct sequences of PCR products amplified from co-cultured PBMCs. [Apetrei (1997)]. Accession numbers Z83284–Z83303.
- 32) **RO3.ID#:** These 24 sequences are from isolates from orphaned Romanian children, ranging in age from 2.5 to 6 years, admitted to a clinic in Tirgu Mures, Romania. All children were referred to this clinic with serious infections and are believed to have been infected horizontally in different orphanages. Virus was isolated after coculture with donor PBMCs. Proviral DNA from cocultured PBMCs was PCR amplified and the PCR products were directly sequenced [Holm-Hansen (1995)]. Accession numbers X77964–X77987.
- 33) **RU1.ID#:** These 3 sequences are from St. Petersburg and Armavir, Russia. Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA [Leitner (1996b)]. RU26 is believed to have been infected by RU20, so only the RU26 sequence is shown here. RU20 was infected by a woman, whose husband was infected by HIV-1 while living in Congo. RU22, RU23 and RU29 were all heterosexually infected by the same HIV-1 infected man, thus only RU22 is presented here. RU30 is believed to have been infected in 1993 while traveling in the Congo. Accession numbers U69655, U69657, U69659, U69661, U69663 and U69664.
- 34) **TW.3341:** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, C, E or G. [Chang (1997)]. Accession number U67765.

G Subtype

At this time there are viral sequences from 73 HIV-1 infected individuals associated with HIV-1 subtype G. The G subtype consensus sequence (G_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

It is noteworthy that all complete genomes containing subtype G regions sequenced to date have been found to be recombinants, having regions that fall into the subtype A clade in phylogenetic analysis. There are at least three circulating recombinant forms that contain regions associated with clades A and G: CRF02(AG) AGI(CY032) and AG(92NG083). See the Nomenclature article in this Compendium for a more in depth discussion of the subtypes and circulating recombinant forms.

- 1) **BE.DRCBL:** This sequence is from a woman living in Belgium who had moved there from the Democratic Republic of Congo (formerly Zaire) in the late 1980s [Debyser (1998)]. The complete genome was sequenced. Accession number AF084936.
- 2) **BE.VII197:** This sequence is from Belgium and was sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. Accession number X96528.
- 3) **BE1.ID#:** This sequence is from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. The sequence VI308 (AJ228227) clusters with AGJ recombinants BFP90 (AF064699, AF057283) and ML84 (AJ245481) and is thus not presented here in subtype G. Accession number AJ228199.
- 4) **BJ1.ID#:** These 2 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. BJ243 is from directly sequenced PCR product, derived via RT-PCR from patient serum RNA. BJ259 is from cloned PCR product, also by RT-PCR from serum RNA [Heyndrickx (1996)]. Accession numbers U61872 and U61874. Subtype A sequences were also determined in this study.
- 5) **CF.4067:** This sequence was associated with the C subtype in first analysis of the C2V3 region [Murphy (1993)], but when a full gp120 sequence became available from this isolate, and phylogenetic analysis was performed including some of the new subtype G sequences, it was more closely associated with G. The full length sequence was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg Cedex, France. It is part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. Accession numbers L11499 and L11500, U43169.
- 6) **CF.15166:** A single sequence from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. It is from PCR clones from cultured proviral DNA [Murphy (1993)]. Accession number L11525. Other sequences from this study were subtypes A, D, E and unclassified. This sequence was originally listed as subtype C in the publication, as was the 4067 isolate.
- 7) **CG1.ID#:** These 5 sequences are from 1988 and 1992 samples collected in Brazzaville and Pointe Noire, Congo [Candotti (1991), Candotti (1999)]. All patients were adults with AIDS at the time of sampling. Subtypes A, D, F, G and recombinants (AG and AE) were identified in this study. Accession numbers for the subtype G samples are AF082306, AF082309, AF082311, AF082316 and AF082320. Gag sequences are also available for some of the samples, with accession numbers M73472–M73480.
- 8) **CG1.CNG30:** This sequence is from Congo (Moriyama, H., unpublished 1998). Accession number AF056186.
- 9) **CI1.ID#:** These 4 sequences are from Abidjan, Ivory Coast. No other information is available (Ellenberger D.L., unpublished 1997). Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000458, AF000460, AF000463, AF000468.
- 10) **CM.276** This sequence is from a 1994–1995 study of 211 Cameroonian AIDS patients [Takehisa (1998)]. Of the 43 HIV isolates sequenced, 17 were subtype A, 1 was subtype B, 2 were subtype C and 1 was subtype G. Accession number AF023072.
- 11) **CM1.ID#:** This subtype G sequence is from one of the samples collected in 1995 from 53 northern Cameroonian Muslim AIDS patients [Mboudjeka (1999)]. Of the 16 HIV patients sequenced, 5 were

- subtype A, 6 were CRF02(AG) IbNG-like, 2 were subtype D, 1 was subtype G and 2 were subtype H. Accession number AF028320.
- 12) **FR1.ID#:** These 2 subtype G sequences are from a study that was done to access the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. There were 2 sequences from patient BCB82, of which only one is presented here. Accession numbers Z95453, Z95461 and Z95463.
 - 13) **GA.LBV217:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a full length env sequence [Janssens (1994b)]. Accession number U09664.
 - 14) **GA1.ID#:** These 2 sequences are from Gabon. G98 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon who moved there from Niger. VI526 is from a 1990 sample from an AIDS patient at the Libreville General Hospital. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Delaporte (1996)]. Accession numbers X90916, G98; X90922, VI526. See also subtypes A, C, D, F and O sequences from this same study.
 - 15) **GB.22:** This sequence is one of three clones from an infected infant in a mother-infant transmission study. The sequences were obtained via PCR from cell lysates, with sequencing of cloned PCR products. The infant was 3 months old at the time of blood drawing, and had pneumonia. [Arnold (1995b)]. Accession numbers U26304–U26306. Envelope sequences for the mother are found in database entries U26301–U26302, and gag sequences for mother and infant are in U26303 and U26307.
 - 16) **GH.NJ3:** This sequence is from Ghana. Subtypes A and D were also detected in this study [Ishikawa (1996)]. Another report from Ghana [Takehisa (1997a)] showed that a Ghana env subtype G isolate was A/G recombinant, with pol gene related to subtype A. Although this sequence appears to be phylogenetically related to that one, we have no pol sequence for this isolate, and thus no solid evidence of it being A/G recombinant. Accession numbers for subtype G is U67683.
 - 17) **GH2.ID#:** These 2 sequences are from Ghana. Subtype A was also detected in this study [Bobkov (1998b)]. Accession numbers for subtype G are AJ225655–AJ225656.
 - 18) **KE1.4089:** This sequence is one of those derived from patients who were part of a study of breastfeeding women from Nairobi, Kenya. Viral DNA was amplified from uncultured patient PBMC, and the envelope V1-V5 region was sequenced after cloning into M13 phage. Other patients from this study had viral subtypes A, C, D and various recombinant forms [Neilson (1999)]. Accession number AF101470.
 - 19) **KP.Kr121** This sequence is from South Korea (Kim, Y.B., et al, unpublished 1995). Accession number X93469.
 - 20) **ML1.ID#:** These 7 subtype G sequences are from a study that looks at the prevalence of different subtypes of HIV-1 and HIV-2 circulating in female sex workers in Bamako, the capital city of Mali [Peeters (1998)]. A total of 176 CSWs were tested and 81 were HIV infected. Of the 81, 63 were infected with HIV-1, 7 were infected with HIV-2 and 11 were dually infected with HIV-1 and HIV-2. HMA assays indicated that 80 percent of HIV-1 infections were with subtype A virus. Only 9 viruses, with ambiguous HMA results, were sequenced. Out of these 9 sequences one was subtype A, one was subtype D and 7 were subtype G. Accession numbers Y14356, Y14357, Y14359, Y14360, Y14361 and Y14363
 - 21) **NG1.ID#:** These 3 sequences are G subtype sequences from Nigeria [Abimiku (1994)]. JP882 and JV832 were derived from AIDS patients, and G3 and G9 from healthy women. G9 was cultured on the T cell line CEM-SS, and the other three isolates were cocultured with uninfected donor PBMCs. DNA from viral cultures was PCR amplified, cloned and sequenced. Accession numbers U13208–U13209, U13211 and U13213. Complete genomes for JV1083 (aka JV832, 92NG083) and G3 (aka 92NG003) are available in database entries U88826 and U88825, respectively. The G3 (92NG003) sequence is subtype A/G recombinant and has been moved to the uncertain section of this analysis.
 - 22) **NG2.ID#:** These 4 sequences are from Nigeria [McCutchan (1999)]. Subtypes A and recombinants were also found. Accession numbers for the G subtypes are AF069935, AF069937, AF069943 and AF069947.
 - 23) **NL.127C** This sequence is one of numerous sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. A sample was collected from the infant at 1.5 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included here [Mulder-Kampinga (1993)] and [Mulder-Kampinga (1995)]. Infant 127

sequences are from Accession numbers Z47817–Z47832. Mother 127 sequences are from entries with Accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in entries with accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see B_NL.114C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.

- 24) **NL1.ID#:** These 4 sequences are from recent immigrants to The Netherlands from Brazil and Democratic Republic of Congo (formerly Zaire). The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. Accession numbers L76880, ZR911976; L76906, LR9401885; L76884, UM9210113; L76902, GH9401230.
- 25) **RU1.ID#:** RU16 is from Volgograd, Russia. The sequence was determined by direct sequencing of PCR product from uncultured PBMC proviral DNA [Leitner (1996b)]. RU16 is a 21 year old female who was infected nosocomially. The nosocomial outbreak in southern Russia began with an index case, an infant born in Elista, to a HIV-1 infected mother whose husband is believed to have been infected in the Congo. Over a period of several months in 1988–1989, over 250 patients, mostly children, were infected parenterally with contaminated needles in several hospitals [Bobkov (1994a), Lukashov (1995), Bobkov (1997c)]. This sequence clusters with the other Russian subtype G sequences, which were also part of this nosocomial outbreak. Accession number U69653. Other sequences from this Russian outbreak can be found with accession numbers L38413, U08355–U08368, U30312, U27445, U51295. Several of these are from the WHO isolate 92RU131 (U27445), from a 3.5 year old female from Rostov-on-Don, isolated in 1992, which has been classified as G/A recombinant [Gao (1996a)]. The CF.4067 sequence is more closely related to these Russian G sequences, than are some other G sequences from Russia, such as BUK3a from Uzbekistan. Other sequences, with only 105 bp of V3 loop data are available in entries with accession numbers U10701–U10859, these include a mother who is thought to have been infected by her infant during breast feeding [Bobkov (1994b)].
- 26) **RU2.ID#:** These two sequences are from Russia and are subtype G in both env and gag [Bobkov (1998a)]. Subtype D/G recombinants were also found, see the RU entry in the unclassified subtype section. Accession numbers for env are AF051508, AF051458 and for gag is AF051509.
- 27) **SE1.SE8552:** This sequence is from a study of 270 HIV-1 infected individuals living in Sweden [Alaue (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 bloods samples were available from 33 men and 42 women, from 15 different African countries. The subtype G sequence was from an individual (SE8552 U76152) who was thought to have been infected in Uganda. Accession numbers for the entire set of all subtypes are U76114–U76186 and L41176–L41179.
- 28) **SE2.ID#:** These sequences are from Sweden [Leitner (1995)]. An env sequence and a complete genome are available for SE6165. Accession numbers L40743–L40745, AF061642.
- 29) **SN1.ID#:** These 3 sequences are from a study done on individuals infected with non-B clade virus who were randomly obtained from a cohort of registered sex workers in Senegal, West Africa [Cao (1997)]. PBMC were separated, cryopreserved and shipped to USA for CTL studies. Of the 14 sequences evaluated 10 were subtype A, three were subtype G and 1 was subtype C. Accession numbers AF020826, AF020830 and AF020831.
- 30) **SN2.ID#:** These 5 subtype G sequences are from Senegal [Kanki (1999)]. Subtypes A, C, D, CRF02(AG) and another AG recombinant were also found in Senegal. Accession numbers for the subtype G sequences are AF085297, AF085301, AF085303, AF085306 and AF085307.
- 31) **TW.267:** This sequence is from directly sequenced PCR product from uncultured PBMCs from Taiwan [Chang (1997)]. Most of the Taiwanese sequences determined in this study were subtype B, but subtypes C, E and F were also found. Database accession number U73058.
- 32) **TW1.ID#:** These 4 subtype G sequences are from Taiwan (Lee, C., unpublished 1997). All 4 are significantly similar to one another, but there is no information provided about epidemiological linkage, so all 4 are presented here. Several sequences were available for each of the 4 patients. Accession numbers AF116739–AF116751.
- 33) **UG.JW3:** This single sequence of subtype G HIV-1 is from a female patient with stage IV disease and CD4 count of 20, from Uganda. The patient had recently migrated to the United Kingdom from Uganda,

but contracted the HIV in Uganda. The sequence is referred to as JW3 in [Kaleebu (1995)] but was previously referred to as K1, by the HIV database. It has been given the name 92UG975.10 by the World Health Organization. Accession numbers U22010, U27426.

- 34) **UZ.BUK3a:** This sequence is from a 39 year old male caucasian heterosexual who lived in Uzbekistan [Bobkov (1996a)]. The patient tested HIV-1 seropositive in June 1991, blood was collected for this sequence in November, 1992. He reported living in Mozambique in 1984-1985 where he was admitted to the hospital several times, but reported having no sexual relations during this time. His wife in Uzbekistan was HIV-1 seropositive in 1991 and died of AIDS in 1995. He reported having sexual relations with a woman in Uzbekistan in 1986-1987 shortly after returning from Mozambique, and this woman and her husband were both found to be HIV-1 seropositive in 1991. Patient BUK died of AIDS in November, 1993. Proviral DNA from uncultured PBMCs was PCR amplified in nested double reactions. The PCR product was cloned into pUC18 plasmid prior to sequencing. Accession numbers U33095, U33096.

H Subtype

At this time there are viral sequences from 11 HIV-1 infected individuals associated with HIV-1 subtype H. The H subtype consensus sequence (H_CONSENSUS) generated from these 11 sequences was based on the most common amino acid found in each position of an alignment. These sequences have been published and/or have been made available for printing in the database by their authors. Eight sequences that are too short for classification are closer to H than to other subtypes.

- 1) **BE.ID#:** These 2 subtype H sequences from Belgium were sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. This VI991 sequence is 98% identical to a sequence called VI1991 with accession number X96532. A complete genome from this isolate is found with accession number AF190127. VI997 has accession number AF190128 and AJ228204.
- 2) **CF.90CF056:** This sequence is from a 1990 blood sample from the Central African Republic. It was originally published as an unclassified subtype, [Murphy (1993)], with accession number L11497. It was later classified as subtype H, [Janssens (1994b)]. The complete genome was sequenced in 1997 (Gao, F., et al, unpublished 1997), and appears to be subtype H throughout the genome, accession number AF005496. L11497 contains an 8 base-pair, frameshifting insertion, relative to AF005496, near the 3' end of its sequence, which is a third copy of a 9 bp direct repeat.
- 3) **CF.4717:** This sequence is from a set of sequences obtained from patients from the Central African Republic [Muller-Trutwin (1999)]. Accession number AF067759.
- 4) **CM.CA13:** A sequence from Cameroon from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence [Janssens (1994b)] and [Nkengasong (1994)]. The H subtype association is not always clearly apparent using some sets of background sequences for comparison, and neighbor joining trees (HIV database, Wouter Janssens, personal communication), although parsimony trees confirmed the original association documented in [Janssens (1994b)]. Accession numbers X80441 and U09667.
- 5) **CM3.ID#:** These 2 subtype H sequences are from samples collected in 1995 from 53 northern Cameroonian Muslim AIDS patients [Mboudjeka (1999)]. Of the 16 HIV patients sequenced, 5 were subtype A, 6 were CRF02(AG) IbNG-like, 2 were subtype D, 1 was subtype G and 2 were subtype H. Accession numbers for subtype H are AF028326 and AF028330.
- 6) **CU.ID#:** These 2 sequences are from Cuba (Gomez, C.E.G.R., unpublished 1997). Accession number Y14415, Y14420.
- 7) **FR1.ID#:** These 2 subtype H sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers Z95457, Z95458 and Z95459.
- 8) **RU.MLY10:** This sequence is from Russia [Bobkov (1996c)]. Patient MLY was a recipient of a 1987 blood transfusion from patient SLH. Patient SLH had a history of sexual contact with an HIV-1 seropositive student from Democratic Republic of Congo (formerly Zaire). Patient MLY gave birth to a child by caesarian section (soon after which she received the transfusion), and the child was later determined to be HIV infected, presumably via transmission during a 3-month period of breast feeding. Accession numbers: U33104 U33105, MLY; U33106 U33107 U33108, SHL.
- 9) **SE1.ID#:** These 2 sequences are from a study of 270 HIV-1 infected individuals living in Sweden [Alaeus (1997)]. 80 of these 270 were thought to have been infected in Africa, prior to moving to Sweden, or infected directly by someone who had just immigrated from Africa. Of these 80, 75 bloods samples were available from 33 men and 42 women, from 15 different African countries. The 2 subtype D sequences were from individuals who were thought to have been infected in Democratic Republic of Congo (formerly Zaire) (SE5930 U76126) and Uganda (SE8646 U76162). Accession numbers for the entire set of all subtypes are U76114-U76186 and L41176-L41179.
- 10) **CD.VI557:** A sequence from Democratic Republic of Congo (formerly Zaire) from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence [Janssens (1994b)]. Accession number U09666.

J Subtype

At this time there are viral sequences from 5 HIV-1 infected individuals associated with HIV-1 subtype J. The J subtype consensus sequence (J_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published or made available to the database for printing.

- 1) **GM1.ID#:** Two sequences from Gambia, as yet unpublished, which cluster with the two Swedish sequences by Leitner in the V3 loop region. Accession numbers U33100, U33102 (Bobkov et al, unpublished 1996). GM4 (U33099, U43105) has been published in [Bobkov (1996b)] as a G/?/C recombinant in the env V1-V5 region, with the ? region covering the V3 loop. See also Gambian sequences of subtypes B and C.
- 2) **SE1.ID#:** Two sequences from Sweden, both from patients who were recent immigrants from Democratic Republic of Congo (formerly Zaire) [Leitner (1995)]. Sample 7022 was collected in December 1993 from an asymptomatic female with a CD4 count of 184. Her first known seropositive sample is from May 1990, but epidemiological investigation indicated that she was infected in Democratic Republic of Congo (formerly Zaire) between 1981 and 1986. Sample 7887 was collected in October 1994 from an asymptomatic male who had tested seronegative in Sweden in January 1993, and who had a seropositive sample in August 1994. His CD4 count was normal at 567. Both patients were heterosexual and a thorough epidemiological investigation revealed no contact or shared contacts between the two. The two sequences were 95% identical to each other over 255 bases of env and 98% identical to each other over 460 bases of gag. Accession numbers L41176 and L41177. Gag gene sequences from these same individuals are in L41178 and L41179 and complete genomes are available with accession numbers AF082394 and AF082395 which were published in [Laukkanen (1999)].

K Subtype

At this time there are viral sequences from 5 HIV-1 infected individuals associated with HIV-1 subtype K. The K subtype consensus sequence (K_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published or made available to the database for printing.

- 1) **BE1.ID#:** These 2 sequences are from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. There are two entries for VI1126 with accession numbers AJ228216 and AF076475, with the latter indicating that patient reported sex with sex workers in the Democratic Republic of the Congo as his risk factor (de Souza et al, unpublished 1999). Accession numbers AJ228142, AJ228216 and AF076475. A gag gene for VI325 is available with accession number L11789.
- 2) **CD.EQTB11:** This sequence is from the Democratic Republic of Congo (formerly Zaire), and 3 sequences from this one isolate are available, a complete genome with accession number AJ249235, and gag and env genes with accession numbers AJ237797 and AJ237808, [Triques (1999)].
- 3) **CM.ID#:** MP535 and MP446 are from Cameroon. A complete genome of MP535 has accession number AJ249239, and gag and env genes from the same isolate AJ237796 and AJ237807. MP446 has only an env gene available with accession number AJ237802 [Triques (1999)]. The MP446 sequence was not long enough on the 5' end to be included here.
- 4) **IL.BC47:** This subtype K sequence is from Israel [Gehring (1997)]. Other subtypes found in Israel in this study were A, B, C, D and CRF02(AG). Accession number for the subtype K sequence is X94386.

Circulating Recombinant form AE

At this time there are viral sequences from 356 HIV-1 infected individuals associated with HIV-1 circulating recombinant form CRF01(AE). The CRF01(AE) subtype consensus sequence (CRF01_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

The AE circulating recombinant form falls within the subtype A clade in phylogenetic analysis of the gag and pol genes and parts of nef and env. In the gp120 region of env, it forms its own clade (E). In the gp41 region of env, it shares similarity with both the A subtype and CRF02(AG) circulating recombinant form. The first complete genome of this AE form to be sequenced was CM240 (accession number U54771). See the Subtype References article by Jean Carr et al in this Compendium (pages III-8 to III-13) for more discussion of the subtypes and recombinant forms.

- 1) **BE1.ID#:** This sequence is from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. See also entries for VI822A and VI822B which are apparently from the same triple-infected patient. Accession number AJ228182.
- 2) **CF.90CF402** 90CF402, previously named CAR-E 4002 or 90CR402, was obtained from a man from Bangui, Central African Republic, who had lymphadenopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. The virus was first adapted to growth in chimpanzee cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lambda cloning and sequencing [Gao (1996b)]. The complete genome is found with Accession number U51188.
- 3) **CF1.ID#:** These 4 sequences are from a set of sequences obtained from patients from the Central African Republic [Muller-Trutwin (1999)]. Accession numbers AF067752, AF067753, AF067766, AF067767.
- 4) **CF1.ID#:** These eight sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA [Murphy (1993)]. Accession numbers L11459–L11460, L11463–L11468, L11476, L11480–L11481, L11504–L11507, L11511–L11513, L11519–L11521 and U43137. Another sequence from patient 4039 is found with accession number U43112.
- 5) **CF2.ID#:** These three sequences were kindly provided prior to publication by Dr. M.P. Kieny of Transgene, Strasbourg Cedex, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C. Mathiot and B. You (Pasteur Inst., Bangui), grown by F. Barre-Sinoussi and A. Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D. Schmitt and M.P. Kieny. Database accession numbers U43110, U43170 and U43173. More sequences from these same isolates are found with accession numbers L11480, L11481, L11506, L11507.
- 6) **CM.CA10:** A single CRF01(AE) sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, CA10 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate [Nkengasong (1994)]. Accession number X80439.
- 7) **CM.CMR10:** This sequence is from Cameroon (Takehisa et al unpublished 1997). Accession number U69991
- 8) **CN1.ID#:** These 15 sequences are from the Guangxi Province, China from blood samples collected in 1996 [Chen (1999)]. Subtypes B, C and D were also identified in this study. Accession numbers AF080192–AF080204, AF080208, AF080213.
- 9) **CZ.CZ4:** This sequence is from the Czech Republic [Quinones-Mateu (1999)]. Accession number AF080135. Other sequences from this set were subtype B.
- 10) **FR1.ID#:** These six sequences are from members of the French military who are believed to have been infected while deployed in Cambodia between 1992 and 1995. Other sequences from this study were subtypes A, B, C, F, and unclassified [Lasky (1997)]. Accession numbers for CRF01(AE) were U58779–U58784.

- 11) **FR2.BCB77:** This CRF01(AE) sequence is from a study that was done to access the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession number Z95454.
- 12) **GB1.11643:** This sequence is from the British isolate 94–11643. The sequence was determined from PCR-amplified lymphocyte DNA. The gag gene from this isolate was subtype A, as is the gag gene from all subtype E virus studied to date. The patient is thought to have contracted the virus in Thailand, but currently lives in the United Kingdom. [Arnold (1995c)]. Accession number U21109.
- 13) **GB2.ID#:** These 9 CRF01(AE) sequences were obtained from a study done on 15 individuals. Eleven of the specimens were from heterosexuals, two were from injecting drug users and one was from a homosexual. Two specimens were from one woman whose risk behavior was not known, and who seemed to be dually infected with subtype B and the CRF01(AE) circulating recombinant form. The specimens were collected in England from individuals whose history indicated that they had become infected in Southeast Asia, particularly Thailand [Belda (1998)]. Accession numbers for the sequences are AJ224176–AJ224189. Gag genes from the same isolates are available with accession numbers AJ224190–AJ224200.
- 14) **ID1.ID#:** These 7 sequences are from a set of 14 sequences from Indonesia [Porter (1997)]. Accession numbers U68188–U68194. Subtype B was also identified in 7 other samples from Indonesia in this study. PCR products were directly sequenced from either uncultured PBMC DNA or cocultured PBMC DNA.
- 15) **IN.ICMCH04:** This sequence is from India (1996 Kalish et al unpublished 1996). Accession number U04536.
- 16) **JP.JH23B:** This sequence is from a Japanese patient dually infected with HIV-1 subtypes B and CRF01(AE) [Xin (1995a)]. Accession number D67089. See also D67090, B_JP.JH23A.
- 17) **JP.NH2:** This sequence is from a study of a Japanese family in which both parents and an infant were infected [Sato (1999)]. The 3 patients were followed over time with 147 env sequences, and also gag sequences. Accession numbers AB014775–AB014873.
- 18) **JP1.ID#:** These three sequences are from one Japanese and two Thai individuals living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells [Weniger (1994)]. NIH3J is from a male Japanese national. NIH2T and NIH4T are from Thai female prostitutes, living in Japan. Accession numbers L32085–L32087.
- 19) **JP2.Pat43:** This sequence is from a study of rapid versus slow progressors in Japan. Patient 43 was a rapid progressor infected through sexual contact in 1992 after testing negative in 1991, [Shioda (1997)]. Other sequences in this study were subtype B. Accession numbers AB002975–AB002987.
- 20) **JP3.ID#:** These 6 sequences are from a study of 26 blood donors from Japan [Kitsutani (1998)]. The other 20 were subtype B. Accession numbers for the entire set are AB012962–AB012994. Accession numbers for these 6 are AB012988–AB012993.
- 21) **KH1.ID#:** These 9 sequences are from HIV-1 infected people from several regions of Cambodia [Kusagawa (1999)]. Accession numbers AB013119–AB013127.
- 22) **KH2.ID#:** These 23 sequences are from Cambodia [Menu (1999)]. The KH009 sequence is an outlier of the CRF01(AE) circulating form, and may be a recombinant between subtypes A and CRF01. More sequence would be needed to determine if it is truly recombinant. Accession numbers AF062011–AF062033.
- 23) **MM.1786:** This sequence is from a dried blood spot collected in 1992 from a female STD patient in Myanmar [Cassol (1996)]. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Accession number U53309.
- 24) **MM2.ID#:** These 3 sequences are from a set of 30 patient samples sequenced from Yangon, Myanmar. Serological subtyping was performed on blood samples from other regions of Myanmar in the same study [Kusagawa (1999)]. IDU in Yangon were all subtype B' as were some of the heterosexually acquired infections. Circulating recombinant form CRF01(AE) was not found in any IDU in Yangon, but only in heterosexual cases. DNA was extracted from PBMCs and PCR amplified. The PCR products were directly sequenced. Accession numbers AB010746, AB010755, AB010765.
- 25) **MY.9214103:** This sequence is from a Thailand infant who was adopted by Malaysians, and now lives in Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced [Brown (1996)]. This report also included subtypes B and C in Malaysia. Accession number U65551.

- 26) **NL.TH94037:** This sequence is from a recent immigrant to The Netherlands from Thailand. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced [Lukashov (1996)]. Accession number L76911.
- 27) **NL.TH94037:** This sequence is from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions. Accession number AF032222.
- 28) **RU1.1164:** This is a one of 3 clones from a single patient RU1164 from Russia [Bobkov (1997b)]. The patient was a 21 year old asymptomatic sex worker who first tested HIV positive in April, 1996. Blood was collected for this sequence in May, 1996. She reported having travelled to Singapore in 1994-1996 and having multiple sexual clients there. DNA from uncultured PBMCs was PCR amplified, cloned and sequenced. Patient RU1164 lives in Khabarovsk kray in the far eastern part of the Russian Federation. This is the first report of CRF01(AE) in Russia. Accession numbers U93607, U93608, U93609.
- 29) **SE.H1** This CRF01(AE) sequence is from a study that includes the analysis of HIV-1 strains in seven cases of mother-to-child transmission in Sweden [Contag (1997)]. HIV-1 isolation from the mothers was conducted during the three trimesters, a few days after delivery and 6 months after parturition. The children were born between June 1989 and August 1992 and were followed by virus isolation at birth and at 3, 6, 12 and 18 months of age unless circumstances called for an extra sample. Breast-feeding was denied by all of the women described. See also A_SE.H4, B_SE5.ID#, C_SE2.ID# and D_SE.H3. Accession Numbers U56263-U56335.
- 30) **SE1.ID#:** This CRF01(AE) sequence is from Sweden. Patient C was a male homosexual, thought to have been infected in Thailand [Karlsson (1999)]. Patient O was a male heterosexual listed as subtype E, but found to be more likely subtype A or recombinant by analysis at LANL (accession number AF014103) and presented in the unclassified subtype section here. Accession numbers for patient C include AF104082-AF104084, AF014420-AF014430 and AF068496-AF068501.
- 31) **SG1.ID#:** These 13 sequences are from Singapore. Other sequences in this set were subtypes A, B and C [Se-Thoe (1998)]. Accession numbers AF004236, AF004237, AF004248, AF004251, AF004252, AF004259, AF004263, AF004265, AF004268, AF004271, AF004273, AF004274 and AF004276
- 32) **TH.93TH253:** This sequence is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93th253 was isolated in 1993 and expanded in human PBMCs, then expanded in H9 cells, and followed by lambda cloning and sequencing. The complete genome has been sequenced [Gao (1996b)]. Accession number U51189.
- 33) **TH.C18:** This sequence is derived from an HIV-1 infected mother enrolled in Bangkok in a perinatal transmission study. Her plasma was screened for gp120 binding antibody, CD4/gp120-binding inhibition, to subtype B (MN/H9), and Thai E (SL7/SupT1) using native gp120 antigens and neutralizing antibody to subtype B and Thai E isolates. In these assays, THC18 plasma showed a pattern of antibody reactivity similar to other E sera. The genetic subtype of this HIV-1 isolate was identified by HMA [Louisirirothanakul (1997)]. Accession number U96782.
- 34) **TH.CM240:** This sequence is from a 21 year old asymptomatic man from northern Thailand. The route of infection is believed to be heterosexual transmission. The blood sample was collected in 1990. The patient's PBMCs were cocultured with stimulated donor PBMCs and proviral DNA was harvested for PCR amplification and sequencing of a cloned full-length genome PCR product [Carr (1996)]. The complete genome is found in the databases with accession number U54771.
- 35) **TH.N764:** This sequence is from a survey of IV drug using prisoners in Thailand. 12 of 13 sequences from Thai prisoners were of subtype B; N764, from patient (THP13) represents the only CRF01(AE) sequence identified in this set, from a prisoner infected in 1989. The sequences were obtained from PCR amplified PBMC DNA [Kalish (1994)]. Accession number U15588.
- 36) **TH.T8178:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was a female prostitute and the sequence was obtained from PCR amplified PBMC DNA [Ou (1993)]. Accession number L19239. See also B_TH.T8174.

- 37) **TH1.ID#:** These twelve sequences are from a set of 23 individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. Referred to as Thai subtype A in [Ou (1992b)] and [Ou (1993)]. (Published erratum appears in *Lancet* **342**:250 (1993).) Accession numbers L07443–L07445, L07447, L07448, L07450 L07457–L07464.
- 38) **TH2.ID#:** Six of these eight sequences are from 16 isolates from HIV seropositive individuals from Thailand. Sequences were from PCR products derived from co-cultured PBMC DNA. The full length envelope gene sequences are available [McCutchan (1992)]. Please note: the “TN-ID#” locus names in the database correspond to the McCutchan et al.’s “CM-ID#” isolates. Accession numbers L03697–L03701 and L03703–L03704. The other two (TH238, TN240) are also from Thailand, DNA from PBMC [Mascola (1994)]. Accession numbers L14571, L14572.
- 39) **TH3.ID#:** Fifteen sequences from asymptomatic individuals from Thailand in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database [De Wolf (1994)]; [Osmanov (1994)]; and [for HIV Isolation (1994)]. Accession numbers U08810–U08811, U08825–U08836, U08742–U08761 and U09131.
- 40) **TH4.ID#:** These three sequences are part of a set of sequences generated for the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. They are clones from expanded culture stocks, and are excised from full gp160 sequences. The sequence ID numbers are abbreviated, for example D3TH966 can be read as DAIDS sequence (D), isolated in 1993 (3), Thai (TH), patient 301966 (966). [Gao (1996a)]. Accession numbers U08456–U08458.
- 41) **TH5.ID#:** These 22 sequences are from twenty two patients with AIDS involved in a study of genotypic and phenotypic characteristics of Thai HIV-1. Blood samples were collected between July and December 1993. All sequences were derived from PCR amplified PBMC DNA, after patient PBMCs were cocultured with virus-free donor PBMCs. CMU01, CMU03, CMU04, CMU05, CMU07, and CMU10 are NSI, the rest are SI, as determined by syncytium formation in the cocultured cells. CM = Chang Mai University Hospital. KH = Kavila Army Hospital. All subjects were males and reported past contact with female sex workers, but no history of drug injection, blood transfusion or homosexual contact. [Yu (1995)]. Accession numbers U25550–U25626. Longer sequences from samples KH003, KH005, KH008, CMU02, CMU08 and CMU010 were determined in 1996 [McCutchan (1996a)]. Accession numbers U48264–U48269.
- 42) **TH6.ID#:** These three sequences are CRF01(AE) sequences from Thailand. Two individuals believed to be dually infected with subtypes B and CRF-AE were analyzed. It is not clear from the paper or the database entries which sequences came from individual 1 and which from 2 [Artenstein (1995)]. Accession numbers U21472, U21474, U21476. See also B_TH5.ID#.
- 43) **TH7.ID#:** These two sequences are from samples collected in 1993 in Thailand. Patients 1018 and 1110 were asymptomatic. [McCutchan (1996a)]. Accession numbers U48273–U48274.
- 44) **TH8.ID#:** These two sequences were from dried blood spots collected in 1992 from a heterosexual (0289) and an IV drug user (0103). DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced [Cassol (1996)]. Accession numbers U53312 and U53313.
- 45) **TH9.ID#:** These 14 sequences are from 84 IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April, 1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced. Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were CRF01(AE) [Kalish (1995), Wasi (1995)]. Accession numbers U22542, U22548, U22553, U22557, U22561, U22567, U22575, U22604, U22609, U22611, U22612, U22617, U22624 and U22625. See also B_TH7.
- 46) **TH10.ID#:** These sequences are from Thailand. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two clones from each isolate were sequenced [Penny (1996)]. Accession numbers U39256 and U39260, 92TH002; U39255 and U39261, 92TH011.
- 47) **TH11.ID#:** These 16 sequences are from northern Thailand. HIV-1 env DNA was PCR amplified from uncultured patient PBMC DNA and PCR products were cloned prior to sequencing. 2 of the patients

each had 3 complete env gene clones sequenced: E11429 and A10121 accession numbers AF015916–AF015921. The others only had the V3 region sequenced. All isolates in this study were CRF01(AE). Patients E11429 and A10121 were both sampled very close to seroconversion. A10121 was a female sex worker who tested HIV antibody negative in April 1993, HIV positive in July 1993, and blood was drawn for these sequences in January, 1994. E11429 was a male military conscript who tested HIV antibody negative in May 1993, seropositive in November 1993, and blood was drawn for these sequences in January, 1994 [Yu (1997)]. Accession numbers for the V3-region set are AF015612–AF015626.

- 48) **TH12.ID#:** These 14 CRF01(AE) sequences are from a set of 95 sequences reported, of which only 26 were submitted to the databases with accession numbers U85085–U85060 [Subbarao (1998)]. The samples of these sequences were collected from 215 asymptomatic HIV-1 individuals from June 1994 through January 1995 at 9 regional medical centres in northern, central and southern Thailand. Out of the 215 participants, 65 were injecting drug users and 150 reported sexual risk behaviors out of which 51 were female sex workers, 41 attended antenatal clinics, 9 had STD's, 41 men with heterosexual behavior and 8 were men who had sex with men. Out of the 215 specimens subtyped, 175 were CRF01_AE, 37 were subtype-B' and 2 were typical subtype-B. See also subtype B for the same study. Accession numbers for the CRF01(AE) sequences are U85060–U85063, U85065, U85067–U85069, U85076–U85078, U85080, U85084, U85085.
- 49) **TH14.ID#:** These 12 sequences are from recent seroconverters in Thailand, sampled in 1998 (McCutchan, F. unpublished 1998). Accession numbers AF070702–AF070713
- 50) **TH15.ID#:** These 91 sequences are from Thailand (Subbarao et al, unpublished 1999). Accession numbers AF081710–AF081780 and AF082536–AF082555.
- 51) **TH16.ID#:** These 31 sequences are from Thailand (Subbarao et al, unpublished 1999). There were 32 CRF01(AE) sequences in this set, but AF151737 and AF151738 were identical, so only AF151737 is presented here. Accession numbers AF151737, AF151739–AF151766, AF151774 and AF151775.
- 52) **TH17.ID#:** These 86 sequences are from Thailand (Auwanit et al, unpublished 1999). There were 86 CRF01(AE) sequences in this set, and no subtype B represented. Accession numbers AB031873–AB032741.
- 53) **TH18.ID#:** These 18 sequences are from Thailand [for HIV Isolation (1994)]. Accession numbers U08743–U08760.
- 54) **TW1.ID#:** These 3 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang (1997)]. Other subtypes found in Taiwan in this study were B, C, F and G. Accession numbers U73060, U73062 and U73070.
- 55) **US.POC30506:** This sequence is from a U.S. serviceman who acquired an HIV-1 infection while deployed in Thailand. He was asymptomatic when the sample for this sequence was collected in 1993 [McCutchan (1996a)]. Accession number U48272.
- 56) **UY1.ID#:** These four sequences are from Uruguayan servicemen who acquired HIV-1 infections while deployed as United Nations peacekeepers in Cambodia in 1993. All four were asymptomatic when samples were collected for these sequences in 1993 [McCutchan (1996a)]. Accession numbers U48275–U48278.
- 57) **VN1.ID#:** These 3 sequences are from South Vietnam [Menu (1996)]. The sequences were from IV drug users in Ho Chi Minh city and Dong Nai, and a female prostitute in Can Tho. A fourth sequence, from Ho Chi Minh city, was found to be subtype B. Accession numbers U29206–U29208.
- 58) **VN2.ID#:** These 4 sequences are from South Vietnam. VN1 and VN2 were from healthy 17 and 25 year old female prostitutes from Can Tho and An Giang. VN3 and VN 4 were from male IV drug users. VN3 was 43, had pruritus and splenomegaly, and was from Nha Trang. VN4 was 31, healthy and was also from Nha Trang [Nerurkar (1996)]. Accession numbers U45239, U45240, U48719 and U48720.
- 59) **VN3.ID#:** These 19 CRF01(AE) sequences are from Vietnam. The blood samples for this study were collected in April/May and August/September 1995 from 8 HIV-1 seropositive CSW in Ho Chi Minh city, Can Tho and An Giang provinces and from 16 IDU in Ho Chi Minh city, Hanoi, Nha Trang and An Giang province. Results showed that CSW and IDU in Vietnam were genetically most similar to CRF-AE strains from Cambodia [Nerurkar (1997)]. Accession numbers U90068–U90090.
- 60) **VN4.ID#:** These 32 sequences are from various cities in northern (NV numbers) and southern (V numbers) Vietnam, sampled in 1998 and 1995 respectively [Kato (1999)]. All patients were asymptomatic, except for NV17, who had AIDS. There were noted differences between strains shared by IVDU and strains

shared by sexual transmissions. Accession numbers AB025050–AB025066 and AB025084–AB025097. Gag sequences for some samples are found with accession numbers AB025067–AB025083.

NOTE:

- 1) While the sequences in this subtype were distinct over this region of env from the other env subtypes, it is not possible to make a distinction between this subtype and subtype A in the gag gene. What this means is that the isolates for which both gag and env are sequenced which cluster together as the “A” subtype in gag, are very distinctive in env and are broken down into two subtypes. env “A” and env “E”. This holds true for the E subtypes sequences that originated in Thailand, as well as the E subtype isolate from the Central African Republic for which gag sequence was obtained. [McCutchan (1992)]; [Louwagie (1993)]; and [Murphy (1993)]. Complete genomes of subtype A and CRF01(AE) viruses became available in late 1996 [Gao (1996b)], genbank accession numbers U51188–U51190.
- 2) The relative lack of diversity in the Thai sequences in this subtype relative to the other subtypes is likely to be a consequence of the short time span of the HIV-1 CRF01(AE) epidemic in Thailand. [McCutchan (1992), Ou (1992b)].

Circulating Recombinant form CRF02_AG

At this time there are viral sequences from 57 HIV-1 infected individuals associated with HIV-1 Circulating Recombinant form CRF02_AG. See the HIV-1 nomenclature article in this compendium for more in depth discussion of the subtypes and recombinant forms. Because this circulating recombinant form is subtype A in the env V3 region, a great many of the V3 region sequences of this form are likely to have been mis-classified as subtype A. A few others may have been noted to be A/G recombinant, without noting that they were of this particular form of A/G recombinant.

- 1) **CI1.ID#:** These 8 CRF02(AG) (IbNG-like) circulating recombinant form sequences from the Ivory Coast were from a set of 13 HIV-1 viral isolates studied in [Janssens (1994a)]. They were initially classified as subtype A, because this region of env in the CRF02(AG) form clusters within subtype A. However, it is clear from their phylogenetic association with the IbNG and DJ263/DJ264 sequences that they belong to the CRF02 form. CI-14 and CI-20 were symptomatic, and the others were asymptomatic. CI-14, CI-45 and CI-47 were serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp sequence. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the one of those clones is presented here. Accession numbers X72024–X72027, X72030–X72036 and X72047–X72061.
- 2) **CM.CAM1725:** This sequence is from Cameroon (Roques et al, unpublished 1998). Accession number AJ006743.
- 3) **CM2.ID#:** These 6 CRF02(AG) sequences are from samples collected in 1995 from 53 northern Cameroonian Muslim AIDS patients [Mboudjeka (1999)]. Of the 16 HIV patients sequenced, 5 were subtype A, 6 were CRF02(AG) IbNG-like, 2 were subtype D, 1 was subtype G and 2 were subtype H. Accession numbers for CRF02(AG) are AF028316, AF028321, AF028322, AF028323, AF028325 and AF028331. Another sequence by these authors, AF119216 which was reported to be from the Democratic Republic of Congo, was more than 98% identical to the sequence here with accession number AF028321.
- 4) **DJ1.ID#:** These three A/G(IbNG) circulating recombinant form sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was 1991. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced [Louwagie (1995)]. Later, complete genomes were PCR amplified and cloned [Carr (1998)] and determined to be AG recombinant with breakpoints similar or identical to those found in IbNG. DJ258 is NSI and uses the CCR5 coreceptor [Trkola (1996b)]. Accession numbers L22939, L22941, L23064 (env); AF063223, AF063224 (complete genomes).
- 5) **FR2.ID#:** These 6 sequences are from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession numbers Z95438, Z95440, Z95443, Z95444, Z95446, Z95451 and Z95452.
- 6) **GA1.ID#:** These 2 sequences are from Gabon. The method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Delaporte (1996)]. Accession numbers X90918, LBV2310; X90921, VI1076. See also subtypes A, C, D, F, G and O sequences from this same study.
- 7) **GH.D687:** A single sequence from an individual from Ghana. Virus was cocultured on PBMCs and the env gene was PCR amplified [Dietrich (1993)]. Accession numbers L07652, X68407.
- 8) **IL.BC47:** These 2 CRF02(AG) sequences are from Israel [Gehring (1997)]. Other subtypes found in Israel in this study were A, B, C, D and K. Accession numbers for the IbNG-like sequences are X94395 and X94396.
- 9) **NG.IBNG:** This sequence is from a complete genome of the Nigerian isolate IbNG [Howard & Rasheed(1996)]. Originally classified as subtype A, it has more recently been shown to be a mosaic between subtypes A and G. Accession numbers U48628 and L39106. It is the prototype of a circulating recombinant form, as many AG recombinants with identical recombination breakpoints which cluster with this sequence in phylogenetic analysis have now been found in many regions of Africa. The entry with accession number U52058 is included here because it seems to be an IbNG-like virus, however the country of origin and other data are not available for this sequence.

- 10) **NG2.ID#:** These 2 sequences are from Nigeria [McCutchan (1999)]. Subtype A and other recombinants were also found. Accession numbers for the CRF02(AG) sequences are AF069933 and AF069941.
- 11) **NL1.ID#:** These 11 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions.
- 12) **RU.IVA6A:** This sequence is from Russia [Bobkov (1996c)]. Patient IVA was infected with CRF02(AG) recombinant virus. Accession number U33103.
- 13) **SE.SE7812:** This subtype A/G sequence is from Sweden (Laukkanen et al, unpublished 1999). The complete genome was sequenced. Accession number AF107770.
- 14) **SN2.ID#:** These 21 sequences are from Senegal [Kanki (1999)]. They were all listed as subtype A in the original entries, but cluster with the CRF02(AG) recombinant sequences within subtype A, and are thus suspected of belonging to this circulation recombinant form. Subtypes A, C, D, G and another AG recombinant were also found in Senegal. Accession numbers for the suspected CRF02_AG sequences are AF085284, AF085288–AF085292, AF085295, AF085296, AF085298, AF085298, AF085300, AF085305, AF085308, AF085311, AF085313, AF085316–AF085318, AF085322, AF085323, AF085325 and AF085327.
- 15) **TW.A3** This sequence is from Taiwan. The two other sequences in this set were subtype A. this sequence was found to be A/G recombinant, clustering with subtype G in the Gag p24 region but with subtype A in env [Lee (1998)]. Patient A3 was a 56 year old female first found to be seropositive in 1991, infected via heterosexual transmission. Accession number AF020602. Gag p24 sequence is found with accession number AF020949.
- 16) **US1.ID#:** These 2 CRF02(AG) sequences are the result of study done from 1992-1994 in which a few newly infected HIV-1 patients who resided in South Bronx New York were studied. Out of these 22 sequences, 2 were CRF02(AG) and 20 were subtype B. This is the first report of a subtype A infection in an American born US resident who had not travelled outside US (patient 866). Patient 912 was born in Africa. Although classified as subtype A in [Irwin (1997)], the sequences cluster with the IBNG sequence which has been shown to be a circulating recombinant form, with subtype A and subtype G regions. Accession numbers U90201, U90202.

Circulating Recombinant form CRF03_AB

At this time there are viral sequences from 41 HIV-1 infected individuals associated with HIV-1 Circulating Recombinant form CRF03_AB. See the HIV-1 nomenclature article in this compendium for more in depth discussion of the subtypes and recombinant forms.

- 1) **LT.LIT173:** This A/B recombinant sequence is subtype B in env and subtype A in gag. The env sequence for the Lithuanian isolates are found with accession number AF082479–AF082482. Gag genes from the Lithuanian samples can be found with accession numbers AF082443, AF082447, AF082448 and AF082449. Other sequences of this recombinant form have been found in the Ukraine, but the largest epidemic seems to be in Kaliningrad Russia [Liitsola (1998)]. The other 3 Lithuanian sequences were subtype B in both gag and env, or subtype A in both gag and env.
- 2) **RU1.ID#:** These 30 A/B recombinant sequences are subtype B in env and subtype A in gag. One complete genome (isolate KAL153) has been sequenced (GenBank accession number AF193276). The env sequences are found with accession numbers AF082450–AF082456, AF082458–AF082478 and AF082485. Gag genes from these recombinants can be found with accession numbers AF082414, AF082397–AF082449, AF082452 and AF082396. Other sequences of this circulating recombinant form have been found in Lithuania and the Ukraine [Liitsola (1998)].
- 3) **RU1.ID#:** These 8 A/B recombinant sequences are subtype B in env and subtype A in gag [Bobkov (1998a)]. The env sequences are found with accession numbers AF051476–AF051479 and AF051483–AF051486. Gag genes from these recombinants can be found with accession numbers AF082414, AF082397–AF082449, AF082452 and AF082396. Other sequences of this circulating recombinant form have been found in Lithuania and the Ukraine [Liitsola (1998)].
- 4) **UA2.NIK81:** This env sequence is from an IV drug user in Nikolaev, Ukraine (Nabatov et al, Unpublished 1998). Accession number AF098952. It was found to be subtype B in env, but is most likely A/B recombinant, based on its similarity to the other CRF03_AB sequences.
- 5) **UA3.UA9:** This subtype A/B sequence is from Ukraine by Grebenjuk et al 1998. Accession number Y16079. Subtypes B and A were also found.

Circulating Recombinant form CRF04-cpx

At this time there are 4 viral sequences from 3 HIV-1 infected individuals associated with HIV-1 Circulating Recombinant form CRF04_cpx. Subsequent to the initial reports of these viruses, which described them first as subtype I, and later as AGI recombinant, they have been further characterized as complex mosaic genomes containing regions of similarity to subtypes A, G, H, and K, as well as some as yet untyped regions [Triques (1999)]. The consensus sequence reflects the sequencing of two clones from one individual and one clone from each of the others. Three of these sequences have been published. See the HIV-1 nomenclature article in this compendium for more in depth discussion of the subtypes and circulating recombinant forms.

- 1) **CY.HOcon:** This is a sequence from one of two individuals who were heterosexual partners of one another, and former IV drug users. They had lived for several years in Athens, Greece as well as in Cyprus. These samples, 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. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from HO32 and one from HO31 were sequenced [Kostrikis (1995)]. Accession numbers U28672, U28673 and U28685. A complete genome from HO32 is available with accession number AF049337.
- 2) **GR1.ID#:** These two sequences are from complete genomes of viruses isolated from individuals living in Greece [Nasioulas (1999)]. PVCH was previously described as GR11 [Nasioulas (1998)]. Accession number AF049292 is also 97PVCH/GR11. The complete genomes are included in the Compendium complete genome alignment.

N Group

At this time there is a sequence from just one HIV-1 infected individual associated with HIV-1 group N that have been published and made available for printing in the database by their authors. The subtypes A–D, F–H, J and K have been grouped together under the heading “M” for main group. “O” group sequences are as different from one another as are sequences from different “M” subtypes, and the N group sequence is nearly equidistant from HIV-1M, HIV-1O and SIV-CPZ sequences. Of particular interest, is the observation that the branching order of HIV-1M, HIV-1N, HIV-1O and SIV-CPZ sequences in phylogenetic trees, is different for different regions of the genome. This has been interpreted as indicating that recombination has occurred between the ancestors of these lineages, probably in the chimpanzee natural host prior to cross-species transmission to humans.

- 1) **CM.YBF30:** This sequence is from 40 year old woman with AIDS living in Cameroon [Simon (1998)]. The authors report that several other patients were infected with HIV-1 that was distantly related to HIV-1 groups M and O, and they found pol gene sequences in two of these individuals which clustered with the N group, separate from the M and O groups. The complete genome is available with accession number AJ006022.

O Group

At this time there are viral sequences from 17 HIV-1 infected individuals associated with HIV-1 group O that have been published and/or have been made available for printing in the database by their authors. The O group consensus sequence (O_CONSENSUS) generated from these sequences was based on the most common amino acid found in each position of the alignment; when there was no consensus in a position an “X” was used. These sequences represent a set of sequences that are extremely divergent relative to other HIV-1’s. The subtypes A–K as well as recombinant forms have been grouped together under the heading “M” for main. “O” group sequences are as different from one another as are sequences from different “M” subtypes, and at some point it may be convenient to classify subtypes of the O group.

- 1) **CM.709:** This sequence is from Cameroon [Takehisa (1999)]. It is from a patient triple infected with an CRF02(AG)-like virus as well as an subtype D and an O group HIV-1. Accession numbers U58154–U58156.
- 2) **CM.ANT70:** The complete viral genome has been sequenced from this viral isolate derived from a symptomatic Cameroonian, CDC stage III. [R. (1990)] and [Vanden Haesevelde (1994)]. Accession number M31171. LTR and partial env sequences were also presented in L20587 and L23119.
- 3) **CM.CA9:** This sequence is from an individual living in Cameroon. [Janssens (1994d)]. Accession number X96522. The pol gene from this isolate has Accession number X78476.
- 4) **CM.MVP5180:** The complete viral genome has been sequenced from an isolate derived from a Cameroonian woman, sampled in 1991; the donor died of AIDS in 1992. The viral isolate MVP-5180 was grown in several human T-cell lines and the monocytic U937 line [Gurtler (1994)]. Accession number L20571. MVP-5180 was shown to be syncytia inducing (SI) on SupT1 cells [Vallejo (1998)].
- 5) **CM.CMR61:** This sequence is one of two O group sequences from a 23 year old female sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa (1997b)]. Accession numbers U58152, U58153.
- 6) **CM2.ID#:** These 10 sequences are O group and mostly pseudogenes, with in/dels that create stop codons (Korber, B.T. et al, Unpublished 1998). Accession numbers AF009025–AF009034.
- 7) **CM2.5267:** This is an O group sequence from Germany (Korber, B.T. et al, Unpublished 1998).
- 8) **ES.1158:** This sequence was from a 35 year old man from Spain. Two blood samples from this same man were collected in April and September 1995. The V3 region was PCR amplified from uncultured PBMCs, cloned into pGEM-5ZF and an individual clone sequenced. The April sequence is shown here. The sequence from the September blood sample (681, Accession number U62617) is also available. Accession number U62618. The tropism of this isolate, and more sequences from this isolate and a related one, were studied by [Vallejo (1998)] with accession numbers AF009608–AF009611. ES1158 and ES1159 were able to infect SupT1 cells but did not induce syncytia.
- 9) **ES.ESP3:** This sequence is from Spain [Mas (1999)]. Accession number AF081812.
- 10) **FR.DUR:** This sequence is from France (Cohen et al, unpublished 1995) Accession number X84327.

- 11) **FR.VAU:** This sequence was derived from an isolate from a French woman who died of AIDS in 1992. DNA was extracted from VAU infected PBMCs, PCR amplified, cloned, and gp160 env was sequenced. The viral isolate was highly cytopathic [Charneau (1994)]. Accession number X80020.
- 12) **FR1.ID#:** These seven sequences are from Cameroonian patients living in France [Loussert-Ajaka (1995)]. PBMC proviral DNA was PCR amplified and 3-6 clones from each patient were sequenced. One of the 3-6 clones is presented. Accession numbers U24562–U24568. Gag gene sequences for these patients are also available with Accession numbers U24706–U24712.
- 13) **GA.VI686:** This sequence is from a 1992 sample from a Gabonese woman with AIDS, taken at the Libreville General Hospital in Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Janssens (1994d), Delaporte (1996)]. Accession number X96526. The pol gene from this isolate has Accession number X78477. See also subtypes A, C, D, F and G sequences from this same study.
- 14) **GQ1.ID#:** These 4 subtype O sequences are from a study in Equatorial Guinea that was selected as a likely region to identify HIV-1 group O infections because it borders Cameroon and includes an island just off the coast of Cameroon [Hunt (1997)]. Four sera were suspected to contain HIV-1 group O related viruses because of their unusual serological reactivity in selected commercial assays and western blots. PHYLIP analysis of the complete env sequences clearly indicated that they clustered with group O sequences and were closest in lineage to HIV-ANT70. Four samples were selected out of which 655Ha was from a sexually transmitted disease clinic, 341Ha was from a pregnant woman, 193Ha was from a tuberculosis patient and 267Ha was from a patient whose diagnosis was unknown. Accession numbers U82990–U82993.
- 15) **US.MD1:** This sequence is from a woman living in the USA who had emigrated from Africa [Vallejo (1998)]. No other patient information was available. MD1 was not able to grow on SupT1 cells and thus did not induce syncytia. Accession numbers AF009612–AF009613.

Unclassified and Recombinant Sequences

At this time there are viral sequences from many HIV-1 infected individuals that are not clearly associated with any of the HIV-1 genetic subtypes A through K or circulating recombinant forms. They either appeared distinct from the subtypes A-K in phylogenetic analysis, or else the subtype association was unclear, with different associations obtained using different tools for analysis, or different regions of the genome. For some of the shorter gene fragments, subtype associations might have been established if more sequence information was available or if a different set of sequences was included in the background set used to define subtype associations. Some of these sequences may be representatives of subtypes as divergent as A-K, but only a single limited sample is yet available. Still others may represent recombinant genomes.

- 1) **AR.20021:** This B/F recombinant sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient was asymptomatic and HIV risk behavior was unknown. Three other samples taken from unrelated patients in 1993 were subtypes F (2) or B (1) [Marquina (1996)]. Accession number U68523.
- 2) **AR.AR8RO:** This sequence is from Rosario Argentina [Fernandez-Medina (1999)]. Other sequences in this study were subtype B (n = 12) or subtype F1 (n = 12). Accession number AF155512.
- 3) **AU.BFP90:** This A/G/J recombinant sequence is from Australia. Accession numbers AF064699, AF057283, AF057284.
- 4) **BE.VI308:** This sequence is from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. The sequence VI308 (AJ228227) clusters with AGJ recombinants BFP90 (AF064699, AF057283) and ML84 (AJ245481). Accession number AJ228227.
- 5) **BE2.VI961:** This subtype D/F1 recombinant sequence from Belgium was sequenced as part of a study analyzing the site specific rates of evolution of the HIV-1 env gene [Van de Peer (1996)]. This sequences came from a Belgian woman whose husband was from the Democratic Republic of Congo. Accession number X96530. A complete genome from this isolate is found with accession number AF076998.
- 6) **BE.VII144:** This sequence is from a study of HIV-1 diversity in Belgium [Heyndrickx (1998)]. Patient VII144 had untyped virus which is possibly C/F recombinant. Accession number AJ228219.
- 7) **BI1.91BU009:** This C/D sequence is from Burundi. It is one of several complete env gene sequences obtained for the World Health Organization. BU009 is from a 36 year old female with CDC stage IV AIDS and pulmonary tuberculosis, from Bujumbura, Burundi. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor-joining tree of the V3 region and with subtype C in other regions [Penny (1996), Ranjbar (1995), Ranjbar(1995)]. Accession numbers L35452–L35459, U39253 and U39254.
- 8) **BR.93BR023** This B/C sequence is part of the WHO Global Programme on AIDS. The virus is NSI and uses the CCR5 coreceptor. Accession number U08779 for env, U86559 for gag and AF009376 for pol gene.
blood donors in Rio de Janeiro Brazil [Tanuri (1999)]. It is subtype D in env V3 and gag, but subtype B in the env gp41 region. One of the 42 sequences reported was subtype B in env and F1 in gag (RJ042). Two were subtype F1 in env and B in gag (RJ063 and RJ069). Thirty-two were subtype B in both env and gag. Six were subtype F1 in env and gag. Accession number AF034020.
- 9) **BR.93BR01904** This B/F1 sequence is part of the WHO Global Programme on AIDS. The virus was derived from a twenty-year old asymptomatic male patient in Rio de Janeiro, Brazil, who presumably contracted the virus through bisexual contact. Blood sample was taken in 1993 [Gao (1996a)]. This sequence appears to be a recombinant of subtypes B and F. Accession numbers U27404, U27408 and U27444. All three confirm the subtype F/B recombination in env.
- 10) **BR.93BR029:** This sequence is another B/F1 recombinant from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced [Penny (1996)]. Another env sequence from this same patient was determined by a second group [Gao (1996a)]. Accession numbers U27413, U39235, U39236 and AF005495 (complete genome). LTR sequence from this isolate is in the entry with accession number U51291.

- 11) **BR.RJ03** This sequence is B/F1 recombinant in the V3 region. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II/A disease in August, 1992, and the PCR product was directly sequenced [Morgado (1994)]. More V3 region sequences from this individual (RJ549 from April 1992) and her sexual partner (RJ548 from April 1992) were also sequenced [Sabino (1994c)]. Accession numbers U00420, U08953–U08955, U08956–U08960, U08962–U08964, U10019–U10026, U08972, U08973 and U08965–U08971.
- 12) **BR1.ID#:** These 3 sequences are from study of asymptomatic blood donors in Rio de Janeiro Brazil [Tanuri (1999)]. One of the 42 sequences reported was subtype B in env and F1 in gag (RJ042). Two were subtype F1 in env and B in gag (RJ063 and RJ069). Thirty-two were subtype B in both env and gag. One was subtype D in env V3 and gag, but subtype B in the env gp41 region. Six were subtype F1 in both env and gag. Accession numbers for the 3 subtype B/F1 are AF034011, AF034014 and AF034017.
- 13) **CD.MAL:** This sequence is from a non-infectious clone of the Democratic Republic of Congo (formerly Zaire) isolate MAL [Alizon (1986)]. The complete genomic sequence and an infectious clone from the isolate MAL are available. MAL is known to be recombinant between subtypes A, D with some unclassified regions formerly referred to as “subtype I”. Accession numbers K03456, A07116 and X04415.
- 14) **CD.Z3:** This sequence is from the 1983 Democratic Republic of Congo (formerly Zaire) isolate Z-3 (non-infectious, possibly due to frame-shift) [Willey (1986)]. Accession number K03347.
- 15) **CD.ZR36:** This sequence is from the Democratic Republic of Congo (formerly Zaire) isolate Z36 [Triques (1999)] Accession number AJ237809.
- 16) **CD.Z321:** This sequence is from the 1976 Democratic Republic of Congo (formerly Zaire) isolate Z321 [Srinivasan (1989)]. Accession number M15896. Earlier listed as subtype A, it was subsequently shown to be recombinant between subtypes A and G [Choi (1997)]. Accession numbers M15896, U76035, U50207, U50208.
- 17) **CF1.ID#:** These four sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. PCR-clones, cell culture, DNA [Murphy (1993)]. Accession numbers L11482–L11483 (4040) are most likely subtype D, L11508–L11510 (4081) are most likely a recombinant, and L11514–L11515 (4087) is most likely subtype A. D. Schmitt provided an unpublished sequence of CF.4081, U43174. Many of these sequences were re-analyzed in [Muller-Trutwin (1999)]. In this re-analysis, a new sequence with accession number AF067755 appeared to be closely related to the 4081 isolate.
- 18) **CF2.ID#:** These 3 sequences are from a set of sequences obtained from patients from the Central African Republic [Muller-Trutwin (1999)]. The 1733 isolate clusters with HO34 and related sequences (accession numbers U28683, U28674, U28719, U28677, U28665). The 11120 and T15 isolates cluster with CF-4081 (accession numbers L11508–L11510). Accession numbers AF067754, AF067755 and AF067761.
- 19) **CG1.ID#:** These 3 sequences are from 1988 samples collected in Pointe Noire, Congo [Candotti (1991), Candotti (1999)]. The patients were adults with AIDS at the time of sampling. Subtypes A, D, F, G and recombinants (AG and EG) were identified in this study. Accession numbers for the recombinant samples are AF082304, AF082313 and AF082317. The EKE isolate is the first recombinant between subtype G and the CRF01(AE) form detected. Gag sequences are also available for some of the samples, with accession numbers M73472–M73480.
- 20) **CM.CA1:** This sequence is 1 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. This sequence was from a symptomatic individual. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate [Nkengasong (1994)]. The other sixteen sequences were subtypes B, E, F, H and A. Accession number X80438.
- 21) **CD.VI191:** This sequence from the Democratic Republic of Congo (formerly Zaire) was from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was 1989. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced [Louwagie (1995)]. Accession number L22952. It was identified as a G/A recombinant sequence based on Gag gene sequence in the entry with accession number L11783.

- 22) **CD1.ID#:** This sequence is subtype undetermined from the Democratic Republic of Congo (formerly Zaire) (Reitz, M. unpublished 1995). Accession number U43101. In some analyses, it seems to cluster with the H or AGI recombinant subtypes.
- 23) **CM.CMR304:** This sequence is from Cameroon [Takehisa (1998)]. Although originally classified as subtype F, and later as subtype F2, it does not seem to cluster with either F1 or F2 sequences in phylogenetic analysis. Accession number U70001.
- 24) **CM.CM61:** This sequence is from Cameroon [Takehisa (1999)]. It is from a patient triple infected with an CRF02(AG)-like virus as well as an subtype D and an O group HIV-1. Accession numbers AF097692–AF097694.
- 25) **CM.G139:** This subtype A/G recombinant sequence is from Cameroon [Delaporte (1996)]. Accession number X96523.
- 26) **CM1.ID#:** These 2 sequences are from Cameroon (Roques et al, unpublished 1998). They might be recombinants with regions of K and/or F2 subtype. Accession numbers AJ006727–AJ006740, AJ006742, AJ006744, AJ006745.
- 27) **CY.HO44-1:** This is a single sequence from two individuals who were heterosexual partners of one another. Patient 16 was a 29 year old bisexual male who was born and lived in Democratic Republic of Congo (formerly Zaire), before moving to Cyprus. He was symptomatic with a CD4 count of 60 and had been seropositive for at least 6 years. Patient 44 was a 32 year old heterosexual female. She was asymptomatic with a CD4 count of 1,136. These samples, like others in this study (see also subtypes A, B, C, and CRF04_cpx) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. They were originally listed as subtype F, because they seemed to be outliers of the F clade. More recent analyses have shown that they are no more closely related to F, than are the subtype K sequences AJ249235 and AJ249239. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. One clone from patient 16 and one from patient 44 were sequenced [Kostrikis (1995)]. Because of the close epidemiological linkage, only the clone from patient 44 is presented here. Accession numbers U28662 (16) and U28679 (44).
- 28) **CY.HO34:** This is a sequence from a set of four individuals who were epidemiologically linked to one another: a father, a mother, their child and a woman who was a heterosexual partner of the father. These samples, like others in this study (see also subtypes B, C, F and I) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR amplified and cloned. Individual clones were sequenced. The father (patient HO34) was 36 years old and asymptomatic. He was known to have been seropositive for at least 2 years and had a CD4 count of 410. The mother (patient HO17) was 35, had been infected for at least two years, had a CD4 count of 2 and died in May 1994. The child (patient HO49) was 2 years old, had been infected since birth, and was asymptomatic, with a CD4 count of 1,211. The partner of the father (patient HO42) was 45 years old, had been infected for at least 4 years, was symptomatic with CD4 count of 80, and died in August 1994. Because of the close epidemiological linkage, only one of the 5 sequences is presented here. [Kostrikis (1995)]. Accession numbers U28683 (child); U28674, U28719 (father); U28677 (father's partner); U28665 (mother). A sequence from the Central African Republic, with accession number AF067754 is reported to be very similar to these sequences of unclassified subtype. The A/C recombinant ZAM184 with accession numbers U86774, U86780 is also roughly 90% identical.
- 29) **ET.3099:** This A/C recombinant Ethiopian sequence is from a complete genome presented in three segments in the database (U92049–U92051). It has been found by the authors to be an A/C recombinant [Sherefa (1998)]. Accession number U92051.
- 30) **FR.BCB69:** This subtype A/C sequence is from a study that was done to assess the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. This sequence is similar enough to the A/C recombinant sequence ZAM184 (Accession U86780) that it is likely to also be recombinant. Accession number Z95445.
- 31) **FR.BCB79** This study presents an env sequence of subtype AGH recombinant virus isolated from a woman living in Paris who moved there from the Democratic Republic of Congo (formerly Zaire). This isolate was identified as an H subtype by a heteroduplex mobility assay on V3-V5 region. The sequence of gag was an A/G recombinant (accession Y13196). Accession number Y13197.

- 32) **FR1.ID#:** This subtype B sequence is from a member of the French military who is believed to have been infected while deployed in Djibouti in 1991. It was classified as subtype C in [Lasky (1997)]. However, it has a GPGR V3 loop tip, and clusters with subtype B in phylogenetic analysis done for this section of the HIV Database Compendium. All other subtype C as of November 1997 have GPGQ at the tip of the V3 loop. Other sequences from this study were subtypes A, B, C, E, and F [Lasky (1997)]. Accession number U58787.
- 33) **FR.CNP1:** This sequence is from a set of 8 sequences from a report that provides molecular evidence for transmission of HIV from an HIV-infected surgeon to one of his patients [Blanchard (1998)]. The orthopedic surgeon was working in a suburb of Paris, France. He probably became infected in 1983 but was not tested for HIV infection until 1994. During those 11 years he had performed 3,004 surgeries, but only one woman, born in 1925, became HIV seropositive after surgery. The sequences from both surgeon and patient are apparently recombinants between subtypes A and F. Accession numbers U85912–U85919.
- 34) **FR.MP84:** This A/E/F1 recombinant sequence is from France [Triques (1999)]. Accession numbers AJ237803. Gag genes are also available.
- 35) **FR.BCB73:** This sequence is from a study that was done to access the reliability of heteroduplex mobility. 199 samples from patients originating from 26 countries and living in France were selected. 37 strains belonging to different subtypes were sequenced and results were found to be in concordance to HMA results [Loussert-Ajaka (1998)]. Accession number Z95449. This sequence might be subtype J or AGJ recombinant.
- 36) **GA1.VI354:** This sequence is from a 1989 sample from a patient with AIDS living in Libreville, Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Delaporte (1996)]. Accession number X90923. See also subtypes A, C, D, A/G and O sequences from this same study. A complete genome of this isolate has been sequenced and has accession number AF076474. A gag gene has accession number L11790.
- 37) **GA.VI525:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced [Louwagie (1994)] and [Janssens (1994b)]. Accession numbers L22953 and U09665. The same isolate was classified as subtype H in gag, [Louwagie (1993)], accession number L11792. The 3' end of the env sequence is also subtype H, while the major portion is subtype G.
- 38) **GA.LBV105:** This sequence is from Gabon. LBV105 is from a 1988 sample from an asymptomatic individual sampled from the general population of Libreville. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced [Delaporte (1996)]. Accession number X90913. See also subtypes A, C, D, F, A/G and O sequences from this same study. LBV10-5 is reported to be subtype AG recombinant in gag and subtype C in env [McCutchan (1996b)]. It is also reported to be recombinant between subtypes A and F in the nef gene, see accession AJ232981, [Jubier-Maurin (1999)].
- 39) **GH.GH8:** This A/G recombinant sequence is from Ghana. Subtypes A and D were also detected in this study [Takehisa (1997a)]. The pol gene for the GH8 isolate has accession number U67039. Accession numbers U67050 and AF056185.
- 40) **GM.GM4:** This sequence is from Gambia and it clusters with the two Swedish sequences by Leitner in the V3 loop region (Bobkov et al, unpublished 1996). GM4 (U33099 U43105) has been published in [Bobkov (1996b)] as a G/?/C recombinant in the env V1-V5 region, with the ? region covering the V3 loop. See also Gambian sequences of subtypes B, C, and J.
- 41) **IN.21301:** This A/C recombinant sequence is from a complete genome from India [Lole (1999)]. The other genomes were found to be pure subtype C. Accession number AF067156.
- 42) **KE.K124:** A Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie (1995)]. This isolate is not clearly associated with D subtype; however, Louwagie and colleagues found that it associated with the D subtype in env, and the A subtype in gag. Using parsimony analysis, we found that it was difficult to determine a clear association,

and this observation was confirmed by Wouter Janssens (personal communication). Accession number L22942.

- 43) **KE1.ID#:** These 7 sequences were derived from patients who were part of a study of breastfeeding women from Nairobi, Kenya. Viral DNA was amplified from uncultured patient PBMC, and the envelope V1-V5 region was sequenced after cloning into M13 phage. Other patients from this study had viral subtypes A, C, D and G [Neilson (1999)]. The subtypes represented in these untyped or recombinant/mosaic isolates are: MM8285, AF101459 C/U; MM6535, AF101467, A/D; MM3055, AF101469, C/D or D/U; MM68, AF101456, U; MM13324, AF101468, A/D; MM13898, AF101457, A/C; MM2760, AF101471, A/G.
- 44) **KR.KR51:** This sequence is from South Korea. Subtypes A, B, C and H were also reported, but the subtype H sequence (KR68), one of the subtype A sequences (KR61) and a subtype C sequence (KR75) were not submitted to the databases [Kim (1999)]. Accession numbers for the entire set are Z92548–Z92668. Subtype A/F accession numbers are Z92620, Z92621, Z92622.
- 45) **LB1.ID#:** These 3 sequences are from a study of 25 sequences from a study of multiple HIV-1 subtypes in Lebanon [Pieniasek (1998)]. Of the 26 samples, 25 were classified as HIV-1: 10 were subtype B, 10 were subtype A, 1 was subtype C, 1 was subtype D and 3 were recombinant or unclassified. The other sample was classified as HIV-2 subtype B. Accession numbers AF025707, AF025713 and AF025715 are HIV-1 subtypes AG, untyped, and AD respectively.
- 46) **ML.95ML84:** These sequences are from Mali (Montavon et al, unpublished 1998) and [Peeters (1998)]. They cluster with BFP90 (Accession number AF057283 and AF064699) and other AGJ recombinants. Accession numbers AJ245481 and Y14358.
- 47) **NG.NG003:** This sequence is from Nigeria [Abimiku (1994)]. Accession numbers U88825 and U13208.
- 48) **NG.NG3670:** This sequence is from Nigeria [McCutchan (1999)]. It clusters with BFP90 (Accession number AF057283 and AF064699) and other AGJ recombinants. Accession number AF069934.
- 49) **NG.NG3678:** This sequence is from Nigeria [McCutchan (1999)]. It is a subtype A outlier and in some regions it clusters with NG1935 (Accession number AF069939). Accession number AF069932.
- 50) **NL.RW94028:** This sequence is from a recent immigrant to The Netherlands from Rwanda. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced [Lukashov (1996)]. Database accession number L76909.
- 51) **NL1.ID#:** This sequence is from a recent immigrant to The Netherlands from the Democratic Republic of Congo (formerly Zaire). The first letter of the ID# represents the country for the previous residence of the patient. The first two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced [Lukashov (1996)]. Although classified as subtype F by Lukashov, the sequence is most closely related to a recombinant genome Gabon with accession number AF076474. Accession L76899.
- 52) **NL2.ID#:** These 5 sequences are from a study in which about 50,000 heterosexual individuals were tested for HIV-1 antibodies in Amsterdam between 1988 and 1996 [Lukashov (1998b)]. 170 individuals were found to be HIV-1 seropositive. Sequences for V3 region were obtained from serum samples of 90 of these individuals. All individuals were AIDS free at the time of sampling. 54 out of these were infected with subtype B virus and none of them originated from sub-saharan Africa. Individuals with non-B viruses originated or had a partner from HIV-endemic regions.
- 53) **NL3.ID#:** These 2 sequences are from BF recombinant genomes found in the Netherlands [Lukashov (1998b)] and [Wolfs (1992a)]. Patients A11 and A12 were a transmission pair so only patient A12 is represented here. Accession numbers L76871, L76901, M91840–M91847 and M91849–M91856.
- 54) **NO.GIL:** This sequence is from a clone of the Norwegian isolate GIL (Jonassen et al, unpublished 1999). The complete genomic sequence is available. GIL is known to be recombinant between a MAL-like strain and subtype H. Accession numbers AJ237565, AJ237571, AJ237574, AJ237578.
- 55) **RU.ID#:** These two sequences are from Russia and seem to be subtype D in gag, but subtype G in env [Bobkov (1998c)]. Although both seem to be D/G recombinants, they do not cluster tightly together in either the D or the G regions and thus may represent two independent D/G recombination events, rather than one circulating form. Accession numbers for env are AF051468, AF051469, AF051471 and AF051472 and for gag are AF051470 and AF051473.

- 56) **RW.92RW009:** This sequence is from a complete genome of a 1992 isolate from Rwanda. The env region clusters with subtype A, but overall the genome is A/C recombinant or mosaic, [for HIV Isolation (1994), Gao (1994a), De Wolf (1994)]. Accession numbers U08631, U08632, U08793, U16220, U16221, U16222, U88823, U13441. A Gag gene is available with accession number U86545.
- 57) **SE.KI4803:** This sequence is from patient number 24 described in [Asjo (1986)] and [Fredriksson (1991)]. Several molecular clones from this patient have been extensively characterized in [Tan (1993)]. Complete env gp120 sequences for 8 clones were determined in [McKeating (1996)] and one of the 8 sequences (clone 13) is presented here. Accession numbers for 7 of the 8 clones (clone 32 was not submitted to the databases) are U57788–U57794. A complete genome from isolate KI4803 has reportedly been sequenced, but not yet submitted to the databases (Fenyo, EM personal communication). The patient had AIDS at the time of viral isolation in 1985. The virus exhibited rapid/high phenotype [Asjo (1986)] and when gp120 from this isolate was swapped into the HXB2 genome, the ability of the chimeras to grow in various cell lines correlated with the gp120 [McKeating (1996)]. Although the amino acid sequences translated from these sequence entries are clearly subtype B, the DNA sequences do not cluster with any of the HIV-1 M-group subtypes. The codon usage of these sequences is similar to the codons used for optimal high-level expression in E.coli, and it is possible that these sequences were reverse-translated from amino acid sequences.
- 58) **SE.O004:** This untyped sequence is from Sweden. Patient O was a male heterosexual listed as subtype E [Karlsson (1999)], but found to be more likely subtype A or recombinant by analysis at LANL. Accession number AF014103.
- 59) **SE1.ID#:** These sequences are from Sweden [Carr (1999)]. The SE6954AD sequence (AF075701) is subtype A/D recombinant with the env region subtype A, and much of the rest of the genome subtype D. SE7108 is mostly subtype A, with a small D-like region. Accession numbers AF014103, AF071473.
- 60) **SE2.ID#:** These A/D recombinant sequences are from Sweden [Leitner (1995)]. Direct sequencing of PCR products with ambiguity codes (R, Y, W etc) for heterogeneous sequence positions. Accession numbers L40748–L40751.
- 61) **SE3.SE8603:** This A/C/D recombinant sequence is from Sweden [Carr (1999)]. The complete genome was sequenced. Accession number AF075702.
- 62) **SE4.SE9488:** This A/C recombinant sequence is from Sweden from a patient most likely infected in Ethiopia [Carr (1999)]. The complete genome was sequenced. Accession numbers AF071474 and U76180.
- 63) **SN.DD900:** This subtype A/G recombinant sequence is from Senegal [Kanki (1999)]. Subtypes A, C, D and G and the CRF02_AG recombinant were also found in Senegal. Accession number AF085314.
- 64) **TZ1.MB6729:** This sequence is part of a set of 86 sequences from samples collected from symptomatic AIDS patients in December 1995 at Mbeya Referral Hospital in southwest Tanzania. Uncultured PBMC DNA was PCR amplified and directly sequenced. Serotyping was also done on all samples to test the ability of serology to subtype these A, C, D and recombinant HIV-1 isolates [Hoelscher (1997), Hoelscher (1998)]. The sequences have not yet entered the databases (12-23-99).
- 65) **TZ2.ID#:** These two subtype AD recombinant sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka (1994)]. Accession numbers U12408, U12409.
- 66) **TZ3.ID#:** These 22 intersubtype recombinant sequences are from Dar es Salaam on the eastern coast of Tanzania [Renjifo (1999)]. Subtypes A, C and D were also found in this study. Accession numbers for the entire set are AF038051–AF038121 and AF106332–AF106472. Accession numbers for sequences that are intersubtype recombinant are AF038061, AF038067–AF038075, AF038094, AF038098, AF038112–AF038117, AF038119–AF038121 and AF106342.
- 67) **TZ4.ID#:** These 2 C/D intersubtype recombinant sequences are from the Mara region of rural northwest Tanzania [Robbins (1996)]. Subtypes A and D were also found in this study. Accession numbers U61881, U61879.
- 68) **UG1.ID#:** Two clones from each of these two D/A recombinant Ugandan isolates were sequenced, but only the sequences of 92UG035 clone 21 and C6080 clone 09 are shown here [Douglas (1996)]. The publication shows C6080 as subtype A, but analyses done at Los Alamos indicate that it is a D/A recombinant, like 92UG035. Other Ugandan isolates sequenced in this study were subtype D. London

subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. Accession numbers U36865, U36866, U36881 and U36883.

- 69) **ZM.ZAM184:** This Zambian sequence is an outlier, though in some phylogenetic analysis it appears most closely associated with the A subtype. In particular it is closely associated with A_CFSAS U43171 (100/100 replicates in parsimony analysis of gp120). Health status of the individual from which the virus was cultured was a woman from Lusaka, Zambia who participated in a clinical research study. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced [Louwagie (1995), Salminen (1997)]. The Salminen paper is based on full-length gag and env genes recovered directly from peripheral blood mononuclear cells or from primary virus cultures, using serial blood samples from a Zambian woman and a sample from her spouse. DNA sequencing and phylogenetic analysis established that two different A/C recombinant forms of HIV-1 predominated at two time points in the woman. A related but distinct recombinant HIV-1 was recovered from her spouse. Intersubtype recombination apparently played a central role in the evolution of HIV-1 in this couple and may contribute substantially to the rapid emergence of HIV-1 variants whenever mixed-subtype HIV-1 infections occur. Accession number L22955. A complete genome from this isolate is now available with accession number U86780 and other clones from this patient and her spouse, from blood samples taken in 1989 and 1990, are found with accession numbers U86768-U86781. Recently, another AC recombinant genome from France has been reported to have the same recombination breakpoints and clusters with this sequence in phylogenetic analysis [Loussert-Ajaka (1998)] accession number Z95445.