Arch Virol (1999) 144: 2291–2311

Archives of Virology

Genetic diversity of HIV-1 group M from Cameroon and Republic of Congo

I. Mboudjeka[1,2], **B. Bikandou**[3], **L. Zekeng**[2], **J. Takehisa**[1], **Y. Harada**[1], **Y. Yamaguchi-Kabata**[1], **Y. Taniguchi**[1], **E. Ido**[1], **L. Kaptue**[2], **P. M'pelle**[3], **H. J. Parra**[3], **M. Ikeda**[4], **M. Hayami**[1], and **T. Miura** [1]

[1]Laboratory of Viral Pathogenesis, Institute for Virus Research, Kyoto University, Kyoto, Japan
[2] Laboratory of Hematology and Immunology, University Center Hospital, Yaounde, Cameroon
[3] Laboratoire National de Santé Publique, Brazzaville, Republic of Congo
[4]Fujirebio Incorporation, Tokyo, Japan

Accepted August 17, 1999

Summary. We analyzed 57 HIV-1 isolates from Cameroon and the Republic of Congo, with respect to the *env* C2V3 and/or the *pol* integrase regions. The results indicated that the topology of the *pol* tree correlated well with that of the *env* tree for four clusters of subtype D, F G and H, suggesting that these trees reflect the true evolution of the overall genome structures of these subtypes. However, of 22 Cameroonian isolates that were classified as subtype A based on *env*, 20 of them diverged in their *pol* sequence into two lineages that were completely different from the prototypical subtype A, tentatively designated as subtypes A1 and A2. The subtype A1 isolates (6 out of 22) were related in their *env* C2V3 regions with prototypical subtype A strain, but in their *pol* regions, they formed an independent cluster that diverged from known HIV-1 subtypes so far reported (except for subtypes I and J). The subtype A2 isolates (14 out of 22), which represent the major epidemic type of HIV-1 in Cameroon, clustered distinctly in both the *env* and *pol* trees with the recently described A/G mosaic strains from Nigeria and Djibouti. These two lineages were not spreading in the neighboring Republic of Congo.

Introduction

Phylogenetic analyses of globally circulating strains of HIV-1 in the envelope protein gp120 region have indentified three distinct groups (M, N and O), and 10 sequence subtypes (A to J) have been proposed within the major group M [18,

24, 26, 29, 31, 39, 58]. The sequence variation among viruses belonging to these different lineages is extensive, with envelope amino acid sequence variation ranging up to 24% between the different subtypes. To explain such diversity, several factors that are known to contribute to the generation of new virus variants and to influence the speed with which these viruses evolve have been enumerated. One is the error-prone nature of the viral reverse transcriptase (RT) which lacks proofreading functions, resulting in nucleotide substitutions, deletions and insertions [9]. A rate of 3.4×10^{-5} misincorporations per replication cycle has been suggested. Another factor is the high rate of virus production (up to 10^{10} virions per day) and the large number of replication cycles that sustain HIV-1 infection in vivo [8, 19, 57].

Recently, recombination has been described as another important factor that might contribute significantly to HIV-1 diversity. Analyses over the last few years have confirmed that substantial numbers of HIV-1 isolates have mosaic genome structures. Some strains that had been previously classified as belonging to one particular subtype based on only *env* region were later revealed as mosaic viruses [13, 27, 28, 40, 46, 54]. Among them, three African strains previously classified as HIV-1 subtype A, are striking. The full-length genomes of these three strains, namely IBNG (from Nigeria) [20, 21] and DJ263 and DJ264 (from Djibouti) [30], were recently revealed to have similar A/G intersubtype recombination patterns [4]. The finding of such hybrid viruses have important implications with respect to viral characterization, the future genetic diversity of HIV-1, and vaccine development. To map and characterize the extent of such inter-subtype recombination, the full-length genome or the sequences of at least two different parts of the HIV-1 genome are needed. Unfortunately, the number of published full-length sequences is very limited because of the time-consuming nature of the sequencing process [34, 40]. The full-length genomes of some HIV-1 subtypes, such as F, G and H, have been recently reported [4, 15] leading to multiple speculations about their origins and dissemination. Whereas the current epidemiology of "subtype E" virus, which is spreading in Thailand and Central Africa Republic (CAR), is now widely believed to be caused by an A/E recombinant virus [5, 14]

The envelope gp120 gene has been particularly targeted for HIV-1 phylogenetic studies and thus is the gene with the greatest number of sequences in the DNA data base [40]. In contrast, the *pol* region of HIV-1 is a relatively poorly investigated part of the viral genome [9, 40, 41]. To date, the classification of various HIV-1 subtypes based on the *pol* phylogeny has been limited to HIV-1 group O and HIV-1 subtypes A, B, C, D, F for group M [40]. For subtypes such as G and H, their equivalent clusters in the *pol* tree have not been elucidated. This is surprising since the *pol* gene also contributes as well to some important biological functions of the virus (production of viral enzyme protease, reverse transcriptase and integrase) and is also known as the principal target of antiretroviral therapy [36, 37, 41]. Some reports have suggested that the V3 loop may be less suitable than other regions of the HIV-1 genome for epidemiological studies, since it appears to be under strong immunology selection pressure after seroconversion

[50, 60, 61]. Consequently, the genetic variation of the V3 region of HIV-1 is too high to be easily interpreted and the analysis of a less variable region such as *gag* or *pol* might offer an alternative region for molecular epidemiological studies of HIV-1 [2, 31]. Because its sequence is highly conserved, the *pol* gene has been successfully used to establish evolutionary relationships among different members of lentiviruses including simian immunodeficiency viruses (SIV) isolated from various non-human primates [7, 12, 16, 35]. Previously we designed a UNIPOL primer set for amplification of the *pol* integrase (*pol*-IN) region and showed that the phylogenetic tree derived from this part of the HIV-1 genome was representative to that derived from the whole *pol* gene [35]. Thus, this *pol*-IN could be useful for clarifying the genetic diversity of HIV-1 and could also provide insights into the true phylogenetic evolution of different HIV-1 lineages of group M.

Since the beginning of the HIV/AIDS pandemic in the early 1980s, the Central African region, especially Cameroon, has become a target for investigating the origin and genetic diversity of HIV-1. The discovery of the divergent HIV-1 group O [18, 29, 40] as well as the discovery of another divergent strain of human retrovirus (human T-cell leukemia virus type 1) [32, 33] were first revealed in Cameroon. A recent report described an HIV-1 strain isolated from a Cameroonian woman that is different from both group M and group O and that phylogenetically falls between HIV-1 group M and SIV cpzGAB [51]. This new strain is being referred to as "group N".

In the present study, we analyzed the sequence diversity of the variable envelope C2V3 region and/or the conservative *pol*-IN region of 57 HIV-1 strains isolated from two Central African countries (Cameroon and Republic of Congo).

Materials and methods

Screening and DNA preparation

Between 1994 and 1997, blood samples were collected in various localities from two Central African countries: Cameroon and Republic of Congo. The individuals from whom the blood samples were obtained were later placed in four groups: AIDS patients, AIDS-related complex patients, tuberculosis patients and asymptomatic carriers. Their plasma samples were initially screened for HIV antibody by particle agglutination (PA) tests (Serodia HIV-1/2; Fujirebio, Tokyo, Japan). The genomic DNAs of seropositive samples were extracted from uncultured peripheral blood mononuclear cells (PBMC) by a glass-milk powder (prep-A Gene DNA purification kit, Bio-Rad). To exclude the possibility of tissue culture artifacts in our sequence analysis, we used directly amplified provirus DNA from PBMC.

Polymerase chain reaction

The varible *env* C2V3 and part of the conserved HIV-1 *pol*-IN region coding for integrase were amplified by nested PCR as previously described [54]. In brief, for the *env* sequence, the outer primer pair M7 (5′-GTC CAA AGG TAT CCT TTG AGC CAA TTC CCA TAC-3′)/M10 (5′-CCA ATT GTC CCT CAT ATC TCC TCC TCC AGG-3′) and the inner pair M3 (5′-GTC AGC ACA GTA CAA TGI ACA CAT GG-3′) M8 (5′-TCC TTG GAT GGG AGG GGC ATA CAT TGC-3′) amplified a fragment of approximately 600 bp (nucleotides

6 567–7 113 in HIV-1 LAI) including the C2, V3, C3, and V4 domains of the *env* gene. For the *pol*-IN region, the outer primer pair HIV-1POL3 (5′-TAA AAG GAG AAG CCA TGC ATG GAC AAG TAGA-3′)HIV-1POL4 (5′-TCA CCT TTC CAG AGG AGC TTT GCT GGT CCT TTC C-3′) and the inner pair UNIPOL1 (5′-AGT GGA TTC ATA GAA GCA GAA GT-3′)/UNIPOL2 (5′-CCC CTA TTC CTC CCC TTC TTT TAA AA-3′) were used to amplify about 300 bp of the highly conserved integrase region corresponding to positions 4 052–4 388 in the HIV-1 LAI sequence. For both *env* and *pol*-IN amplification, the first round of PCR was carried out in 50 μl of PCR mixture containing 1 μg of chromosomal DNA, 50 mM KCl, 10 mM Tris (pH 8.3), 6 mM $MgCl_2$, 1 mM dithiothreitol, 20 μg/ml bovine serum albumin, 200 ng of each primer, 0.2 mM of each dNTP and 2.5 U of Tth DNA polymerase (Toyobo Co., Japan). The amplification was performed in a Thermal Sequencer (Iwaki Glass Co., Japan) for 35 cycles [30 sec at 94 °C, 30 sec at 45 °C for *pol*-IN PCR (at 55 °C for *env* PCR) and 1 min at 72 °C] with an initial denaturation step for 5 min at 94 °C and a final extension for 5 min at 72 °C. The second round of PCR used 5 μl of the first PCR product as a template with the same cycling parameters except for the annealing temperature (50 °C for *pol*-IN and 60 °C for *env* PCR).

Molecular cloning and DNA sequencing

The amplified DNA fragments were fractionated by agarose gel electrophoresis, purified with a Geneclean II kit (Bio 101, Inc., USA), blunt-ended by Klenow enzyme and ligated into the Sma I site of pUC119. After transformation of competent *Escherichia coli* (DH5 alpha), the purified plasmid clones were sequenced by the dideoxy chain termination method using an automated DNA sequencer (373A; Applied Biosystems, USA). For each positive sample, three to four plasmid clones were sequenced and the consensus sequences were used for the phylogenetic analysis.

Sequence alignment and phylogenetic analyses

The phylogenetic relationships of the newly derived Central African viruses were estimated from comparisons of their sequences with previously reported sequences that are representative of HIV-1 group M. Multiple *pol*-IN and *env* sequence alignments were obtained from the Los Alamos sequence database (http://hiv-web.lanl.gov/HTML/alignments.html). The references used are those of full-length representatives of different HIV-1 subtypes whose sequences overlapped those of our *pol*-IN and *env* (C2V3) regions [40]. The reference strains included both non-recombinant and recombinant strains. The non-recombinant strains included subtype A (92UG037), subtype B (LAI, HXB2R, OYI, MN), subtype C (C2220, D747, D1044), subtype D (NDK, ELI), subtype F (93BR020) and subtype H (90CF056). The recombinant references included A/E (90CR402, CM240, 93TH253), A/D(MAL) as well as the recently described A/G mosaic strains (IBNG, DJ263 and DJ264) [4]. For subtype G, some reference strains (92NG003, 92NG083, SE6165 and HH8793) were recently published [4, 15] but it is still unclear whether these strains are really authentic non-recombinant subtype G. Each nucleotide sequence was aligned by using the Computer program CLUSTAL W [55]. The pairwise evolutionary distances were estimated by Kimura's two-parameter method [23]. Phylogenetic trees were constructed by using the Neighbor-Joining method [48] and the confidence values for individual branches of the resulting trees were evaluated by bootstrap analysis using 100 replications [11]. The trees were then visualized by TREEVIEW [43].

The nucleotide sequence data reported in this paper will appear in the GenBank nucleotide sequence database with the following accession numbers: AF028316-AF028332, AF028074-AF028090, and AF120186-AF120205.

Results

Subjects and specimens

The isolates found to belong to subtype Ds (2 Cameroonian and 3 Congolese), subtype Fs (3 Cameroonian), subtype Gs (2 Cameroonian and 7 Congolese), subtype Hs (2 Cameroonian and 5 Congolese) and some of the subtype As (3 Congolese and 22 Comeroonian) in *env* C2V3 analysis were further analyzed in the *pol*-IN (Table 1). Of the 57 HIV-1 isolates obtained in this study, 49 were analyzed in both the *env* and *pol*-IN regions and the other 8 (prototypical subtype A from Congo) were analyzed in the *env* region only. As shown in Table 1, the samples belonged to either patients with AIDS, AIDS-related complex, tuberculosis or people who were asymptomatic at the time of blood collection.

Phylogenetic analyses based on the env and pol-IN regions

Our *pol*-IN phylogenetic tree depicts the overall topology obtained with the *env* C2V3 gene analysis (Figs. 1 and 2). The topologies of the phylogenetic trees containing the various HIV-1 reference strains (A, B, C, D, E, F, G and H) based on portions of the *env* and *pol*-IN regions generally correlated each other. Of the 49 samples analyzed in this study, based on their positions in the trees, 5 were classified as subtype D, 3 were classified as subtype F, 9 were classified as subtype G, 7 were classified as subtype H (Table 1). In addition, of the 22 Cameroonian isolates presumed to be subtype A, only 2 isolates clustered with prototypical subtype A. The remaining 20 isolates clearly diverged in two disinct lineages as shown by the two phylogenetic trees. We proposed to call these two lineages subtypes A1 and A2.

Two dominant HIV-1 subtypes in Cameroon

Two dominant HIV-1 subtypes to have spread in Cameroon was found in this study. One of the origin of these viruses has been clarified with the help of a recent report revealing the structure of three presumed *env* subtype A strains (IBNG, DJ263 and DJ264) [4]. The first one (IBNG) was from Nigeria whereas DJ263 and DJ264 were isolated from French Foreign Legion soldiers assigned to peacekeeping in Djibouti in the late 1980s. The full-length genomes of these three strains were shown to have very similar patterns of A/G mosaic structure, and it has been tentatively suggested that these three isolates be called “subtype IBNG”, referring to their possible Nigerian origin. However, we made the surprising observation that the majority of our Cameroonian HIV-1 isolates tightly clustered in *env* and *pol*-IN phylogenetic trees with the above three A/G references (IBNG, DJ263, DJ264) (Figs. 1 and 2). As shown in Table 1, within the 22 Cameroonian *env*-related subtype As, only two isolate (94CM323 and 95CM897) could be considered as prototypical subtype A viruses. The remaining 20 isolates were clearly divided into two divergent groups as shown in the trees. The larger group (subtype A2), consisting of 14 out of 22 isolates, clustered distinctly together with the above A/G reference strains in both the *env* and *pol*-IN trees suggesting that they

Table 1. Epidemiological, clinical data and genotype of 57 Central African HIV-1-infected patients for whom the *env* C2V3 or *pol* integrase regions were analyzed

Sample[a]	Sex/age[b]	Location	Disease status[c]	HIV-1 genotyping[d]	
				env C2V3	pol-IN
94CM61	F/22	Awaé	ARC	A2	A2
94CM90	M/34	Batouri	AS	A2	A2
94CM116	F/28	Batouri	AS	A1	A1
94CM142	M/24	Batouri	AS	A1	A1
94CM153	M/35	Batouri	AIDS	A2	A2
94CM154	F/40	Batouri	AS	A2	A2
94CM157	M/49	Batouri	AS	A2	A2
94CM158	M/34	Sangmélima	AIDS	F	F
94CM276	NA/NA[e]	Douala	NA	G	G
94CM278	NA/NA	Douala	NA	A2	A2
94CM279	NA/NA	Douala	NA	A2	A2
94CM304	M/NA	Douala	ARC	F	F
94CM323	NA/NA	Douala	NA	A	A
95CM371	M/50	Douala	AIDS	A2	A2
95CM419	F/44	Douala	ARC	A2	A2
95CM653	F/NA	Douala	AS	A2	A2
95CM659	M/31	Douala	AS	F	F
95CM661	M/22	Douala	AS	D	D
95CM844	F/40	Ngaoundéré	AIDS	A1	A1
95CM847	M/20	Ngaoundéré	ARC	G	G
95CM851	F/31	Ngaoundéré	TB	A2	A2
95CM862	M/29	Ngaoundéré	ARC	A2	A1
95CM864	M/26	Ngaoundéré	AIDS	A1	A2
95CM873	M/25	Ngaoundéré	ARC	A2	A2
95CM874	M/32	Ngaoundéré	ARC	H	H
95CM884	F/33	Ngaoundéré	TB	A1	A1
95CM890	F/35	Ngaoundéré	TB	A1	A1
95CM891	F/25	Ngaoundéré	AIDS	D	D
95CM894	M/31	Ngaoundéré	ARC	H	H
95CM895	M/32	Ngaoundéré	AIDS	A2	A2
95CM897	F/31	Ngaoundéré	ARC	A	A
96CG12	M/40	Brazzaville	AS	A	A
96CG13	M/30	Brazzaville	AS	G	G
96CG14	M/35	Brazzaville	ARC	D	D
96CG16	M/30	Brazzaville	ARC	H	H
96CG17	M/35	Pointe-Noire	AS	A	ND
96CG21	M/33	Pointe-Noire	AS	A	ND
96CG22	M/38	Pointe-Noire	AS	G	G
96CG23	M/32	Pointe-Noire	AS	H	H
96CG24	M/25	Pointe-Noire	AIDS	A	ND
96CG27	F/25	Pointe-Noire	ARC	H	H
96CG30	M/25	Pointe-Noire	ARC	G	G
96CG38	M/32	Pointe-Noire	AS	G	G
97CG152	F/40	Brazzaville	AS	A	ND
97CG246	NA/NA	Brazzaville	AS	D	D

Table 1 (*continued*)

97CG250	NA/NA	Brazzaville	AIDS	D	D
97CG252	NA/NA	Brazzaville	AS	G	G
97CG257	F/36	Brazzaville	AS	A	ND
97CG276	F/32	Pointe-Noire	AIDS	A	A
97CG281	F/48	Pointe-Noire	AS	A	ND
97CG282	F/41	Pointe-Noire	ARC	A	A
97CG284	F/38	Pointe-Noire	AIDS	A	ND
97CG307	F/20	Pointe-Noire	ARC	A	ND
97CG313	M/35	Pointe-Noire	ARC	G	G
97CG314	M/44	Pointe-Noire	AS	H	H
97CG315	F/43	Pointe-Noire	AS	G	G
97CG316	F/33	Pointe-Noire	AIDS	H	H

[a]The isolate names are preceded by the year of collection. *CM*, *CG* Camerroon and Republic of Congo, respectively

[b]*F* Female; *M* male

[c]*AS* Asymptomatic carrier; *ARC* AIDS-related complex; *TB* tuberculosis

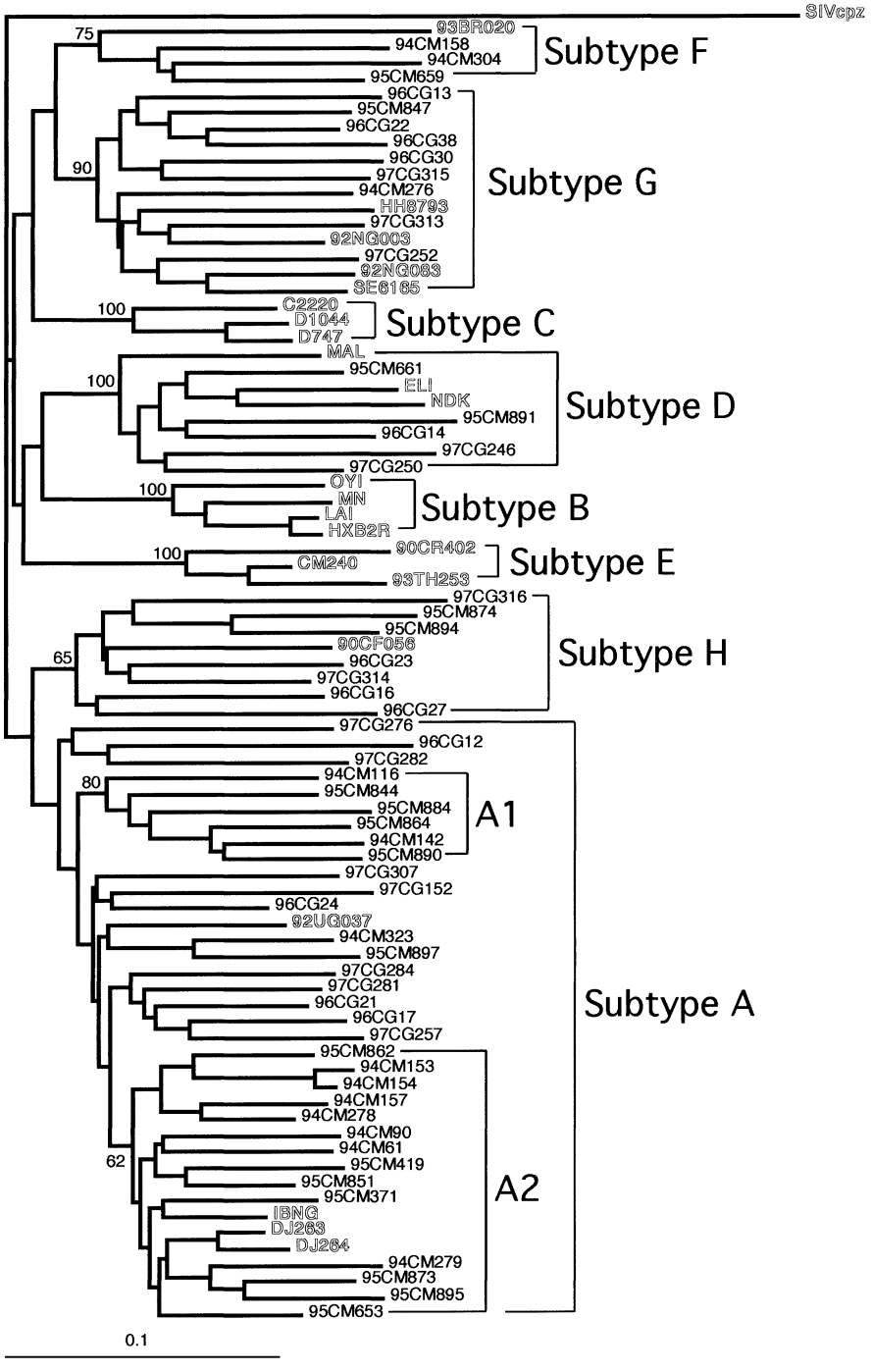
[d]*ND* Not done

[e]*NA* Data not available

are also mosaic viruses. The cluster in the *pol*-IN phylogenetic tree was remarkable in its close relationship with the subtype G cluster, which clearly indicates that all of these A2 viruses shared the subtype G sequence in this region (Fig. 2). The average nucleotide divergence of this group of viruses was around 11% in the *env* sequence and only 2% in the *pol* sequence.

On the other hand, another group comprising 6 out of 22 presumed Cameroonian subtype A isolates was also identified. This group, which we temporarily designated as subtype A1, remained phylogenetically related in their *env* C2V3 region to the reference sequence of prototypical subtype A (92UG037) (Fig. 1). In contrast, in their *pol*-IN sequence, they formed an independent and unique cluster, different from any of the known HIV-1 *pol* subtypes reported so far although it is completely included to group M (Fig. 2). It is unclear whether A1 cluster belong to subtype I or J because the reference sequences of these subtypes in the *pol* region are not available. At least, however, this cluster in the *pol*-IN phylogeny showed no evidence of a phylogenetic link with the subtype A or A2 clusters. The average nucleotide divergence was around 12.5% in the *env* sequence analysis and 2.4% in the *pol* sequence analysis.

These two groups appeared to be located in Cameroon. No Congolese isolate clustered with subtype A1 and A2 cluster. The Congolese isolates rather grouped in this C2V3 region with the known prototypical subtype A references whereas the Cameroonian isolates branched into two distinct types that diverged from subtype A. These results indicated that out of 22 Cameroonian *env*-related subtype A, only 2 (9.1%) of the isolates belonged to prototypical sybtype A, 6 (27.3%) belonged to subtype A1 and 14 (63.6%) belonged to the new subtype A2. In contrast, all of the 11 Congolese isolates that were classified as subtype A by *env*, may belong to prototypical subtype A.


SIVcpz
93BR020
94CM158
94CM304
95CM659
Subtype F
96CG13
95CM847
96CG22
96CG38
96CG30
97CG315
94CM276
HH8793
97CG313
92NG003
97CG252
92NG083
SE6165
Subtype G
C2220
D1044
D747
Subtype C
MAL
95CM661
ELI
NDK
95CM891
96CG14
97CG246
97CG250
Subtype D
OYI
MN
LAI
HXB2R
Subtype B
90CR402
CM240
93TH253
Subtype E
97CG316
95CM874
95CM894
90CF056
96CG23
97CG314
96CG16
96CG27
Subtype H
97CG276
96CG12
97CG282
94CM116
95CM844
95CM884
95CM864
94CM142
95CM890
A1
97CG307
97CG152
96CG24
92UG037
94CM323
95CM897
97CG284
97CG281
96CG21
96CG17
97CG257
95CM862
94CM153
94CM154
94CM157
94CM278
94CM90
94CM61
95CM419
95CM851
95CM371
IBNG
DJ263
DJ264
94CM279
95CM873
95CM895
95CM653
A2
Subtype A
75
90
100
100
100
65
80
62
0.1

Two phylogenetic trees constructed by other methods (the Maximum-parsimony method and the Maximum-likelihood method) gave similar results (data not shown). It is interesting that two of the Cameroonian isolates (95CM862 and 95CM864) permuted their positions in the trees, which raises the possibility that recombination occurred between these two lineages of virus (Table 1, Figs. 1 and 2). Also noteworthy is the difference between the consensus subtype A, and the two Cameroonian subtypes (A1 and A2) at the C2V3 amino acid level as shown by the identity of some conserved residues (Fig. 3). Especially at the apex of the V3 loop, all the subtype A1 isolates were characterized by the hexameric amino acid sequence GPGQAF, whereas all of the subtype A2 isolates have the sequence motif GPGQTF.

HIV-1 subtypes D and F

As shown in Figs. 1 and 2, the Cameroonian and Congolese subtype D isolates clustered together with the subtype D references (ELI and NDK from former Zaire). However, they branched earlier than ELI and NDK in both trees with the nucleotide divergence ranging between 18.9% and 23.1% in *env* sequence and between 7.6% and 11.7% in *pol*-IN sequence. Two of these isolates had a tetrameric GPGR in their C2V3 amino acid sequence and one of them (96CG14) had 35 amino acids (instead of 33) in the V3 loop and a replacement of residue Y by F in the CTL recognition site (Fig. 3). Three isolates of subtype F of Cameroon origin also consistently clustered with the unique Brazilian F reference (93BR020), although they were slightly distant from the Brazilian subtype F in the *pol*-IN tree as well as in the *env* tree.

HIV-1 subtype G

Also interesting in this study are the *pol*-IN results of the 9 Central African isolates (2 Cameroonian and 7 Congolese) that were found to have *env* genotype G. Recently, two independent studies have reported the full-length genomes of four African strains that were previously classified as sybtype G based on the *env* region [4, 15]. Three (92NG083, SE6165 and HH8793) of them were described as probably prototypical (non-recombinant) G viruses and one (92NG003) was described as having an unusual A/G recombination pattern (with some part of the genome unclassified). Two of these four isolates (92NG003, 92NG083) originated

Fig. 1. The *env* C2V3 phylogenetic relationship of 57 Central Africa HIV-1 isolates. Reference sequences for all known group M subtypes were obtained from the Los Alamos sequence database [42] with highlighted letters including those of subtype F (93BR020 from Brazil), H (90CF056 from Central Africa Republic), G (92NG003 and 92NG083 from Nigeria, SE6165 from Kenya and HH8793 from Congo), and A/G (IBNG from Nigeria, DJ263 and DJ264 from Djibouti) reported recently and which are the representative full-length sequences of their respective clades [6, 14]. The tree was constructred by the neighbor-joining method [49] and the reliability of the topologies was estimated by performing a bootstrap analysis [12]. Horizontal branch lengths are drawn to scale and vertical separation is for clarity only. The tree was rooted with SIVcpzGAB

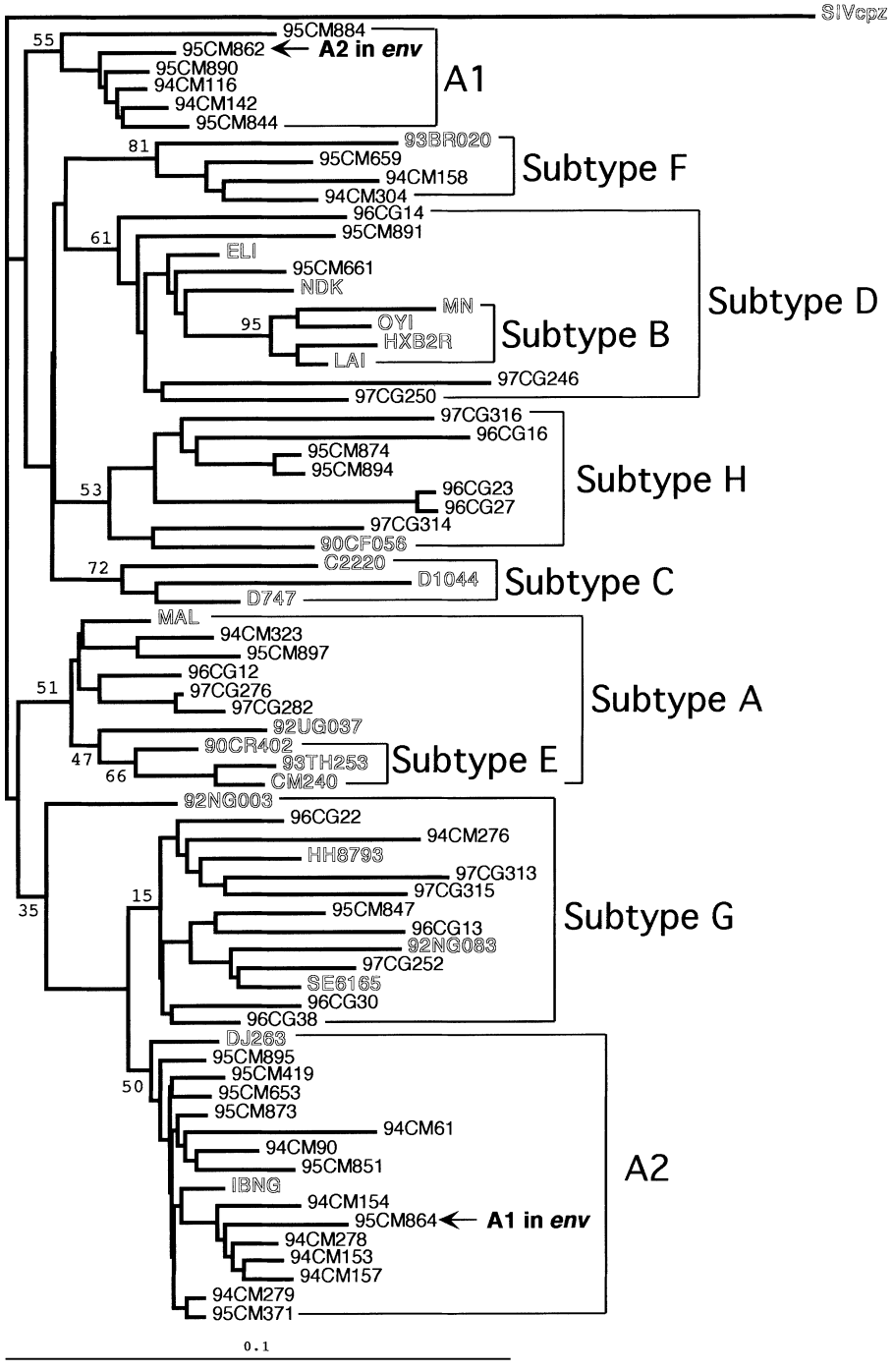


Fig. 2. The *pol* integrase (*pol*-IN) phylogenetic relationship of the newly characterized 49 Central African HIV-1 isolates. The tree was constructed by neighbor-joining method [49] using 288 bp encompassing the integrase region. Labelling of a subtype for each strain is done on the basis of *env* analysis. The tree was rooted with SIVcpzANT. See legend of Fig. 1 for further details

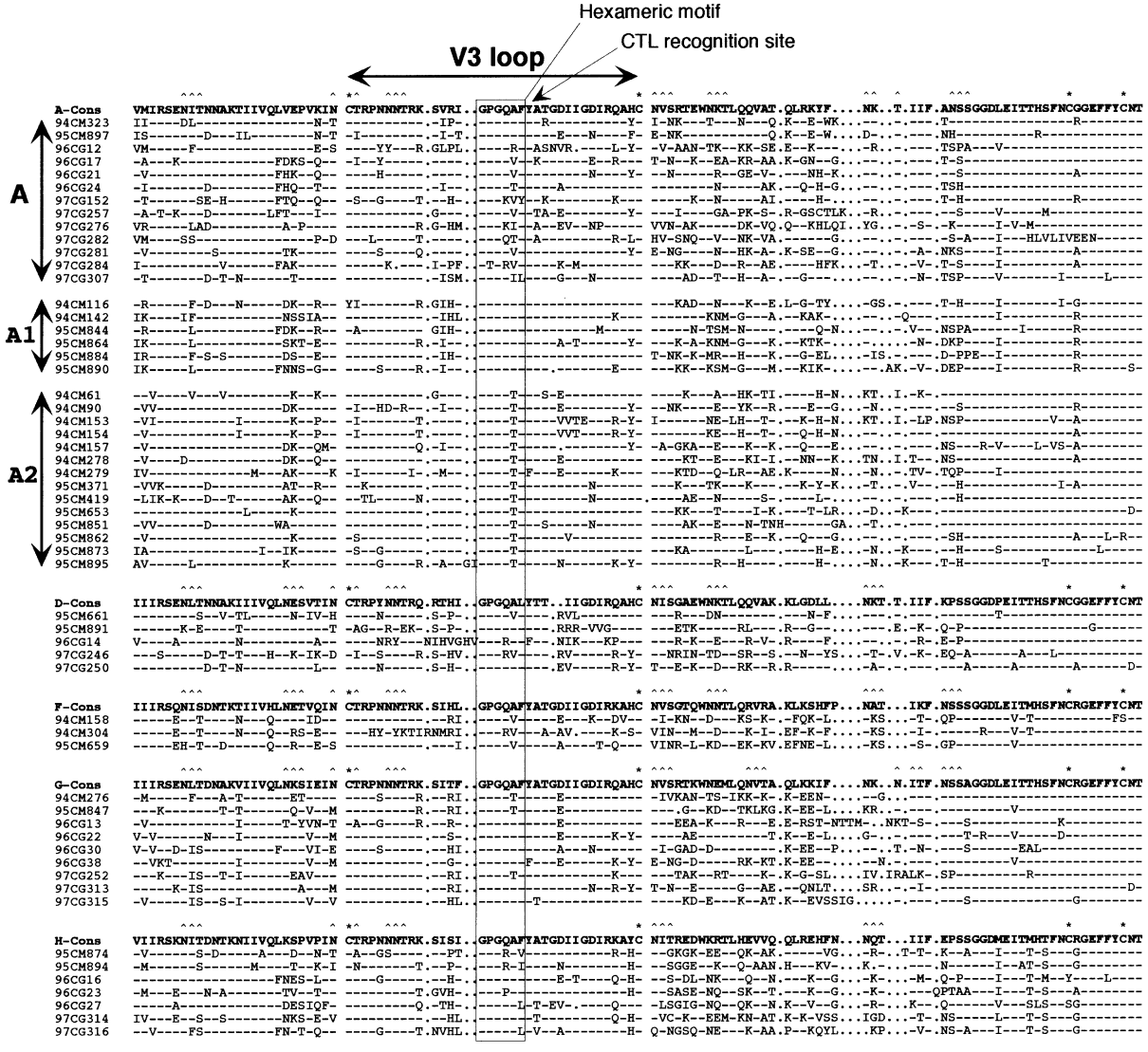


Fig. 3. Deduced amino acid sequences of the C2V3 region of 59 Central African HIV-1 isolates. The most recent compiled consensus amino acid sequence alignment from Los Alamos [42] is given on the top line using the single-letter code. Dashes indicate concurrence with the top sequence in the alignment, period a deletion. The carets show where the potential N-linked glycosylation sites are found in the consensus

Table 2. Total prevarence of HIV-1 subtypes based on env C2V3 sequence in Cameroon and Republic of Congo

	Total	A	A1[a]	A2	D	F	G	H
Cameroon	31	2(6.5)[b]	6(19.4)	14(45.2)	2(6.5)	3(9.7)	2(6.5)	2(6.5)
Congo	26	11(42.3)	0(0.0)	0(0.0)	3(11.5)	0(0.0)	7(26.9)	5(19.2)

[a] A1, A2; tentatively named subtype in this study
[b] % in parenthesis

from Nigeria [15] and the other two were isolated in Sweden (SE6165) and Finland (HH8793) but presumably originated in Congo and Kenya, respectively [4]. We used these four strains as references for our subtype G analyses. Remarkably, as seen in the trees (Fig. 1 and 2), all 9 isolates of *env* subtype G grouped distinctly in a unique cluster in the *pol*-IN tree together with those of the G references that were described as non-recombinant subtype G. In contrast, 92NG003 was easily identifiable in this *pol*-IN tree analysis as occupying an intermediary position between the subtype A and subtype G clusters (Fig. 2).

HIV-1 subtype H

All of the 7 new isolates of subtype H clustered together in the *pol*-IN and *env* phylogenetic trees with the recently published and unique full-length subtype H (90CF056 from Central Africa Republic) [15]. This cluster of subtype H, which has not yet been described in the *pol* phylogeny, probably reflects the true overall genome structure of HIV-1 subtype H. The Central African subtype H isolates were relatively more divergent than those of other subtypes, showing long branches in the phylogenetic trees (Figs. 1 and 2). They diverged at the nucleotide level ranging from 13.4% to 25% in the *env* region and from 0.7% to 12.1% in the *pol*-IN region. As shown in Fig. 3, the two Cameroonian H isolates have GPGR, one of the Congolese isolates has GPGP and the other Congolese isolate has the consensus GPGQ in the V3 tip. It should be noted that the *env* subtype H consensus sequence used in this study was obtained from the DNA database (as published in 1996) [40], which is based on only two subtype H strains from Africa (CA13 from Cameroon and VI557 from Zaire) [42].

Pol-IN nucleotide alignment studies

The naturally occurring G-to-A hypermutation has been shown to frequently occur among lentiviruses [6, 17, 56]. To exclude this possibility as a simple explanation for the new clustering in our *pol*-IN tree, we investigated the nature of the nucleotide substitutions involved in the *pol* integrase sequence of the two dominant subtypes in Cameroon. We also hoped to gain some insights at the nucleotide level concerning the relationship between subtype A and subtypes A1, A2 and G in the *pol*-IN phylogeny (Fig. 2). All of the A1 (6 isolates), A2 (14 isolates), prototype A (5 isolates) and subtype G (9 isolates) were aligned with the recent *pol*-IN consensus-A sequence by using the CLUSTAL W program [55]. As shown in Fig. 4, we mapped 29 critical positions within this conserved *pol* integrase region that determines the *pol*-IN phylogenetic position of different clusters. The nature of the common substitutions found in each group clearly excluded the hypothesis of G-to-A hypermutation and reinforced the idea that the different groups rather envolved as an independent lineage. Especially for subtype A1, the nucleotide substitutions at several positions (as compared with the consensus-A alignment) was noteworthy. For example the substitutions at positions 1 (G to A), 5 (C to T) and 6 (? to C), are those found in subtypes B, D and HIV-1 group O, whereas the substitution at position 19 (A to G) was unique

in HIV-1. On the other hand, subtype A2 has many common substitutions shared with subtype G (nucleotide T at positions 2, 6 and 27, nucleotide G at positions 8, 26 and 29, nucleotide A at positions 7, 11, 15 and 22, and nucleotide C at positions 14, 18 and 25) although it has unique substitutions (nucleotide G at positions 4 and 12).

Discussion

Integrity of pol-IN and env-based analyses

HIV-1s other than subtype B also cause the vast majority of new HIV-1 infections worldwide, but they are only infrequently studied and well-characterized full-length reference sequences are still needed [40]. This lack of reference sequences also confounds the indentification and analysis of mosaic genomes which appear to be arising with increasing frequency in areas where multiple sequence subtypes are co-circulationg. Without detailed analyses, genetic recombination between subtypes may obscure phylogenetic relationships between sequences. Therefore, we must take care for over-interpreting the subtype designations based on a single part of the HIV-1 genome. In this study we demonstrated that the analyses of the *pol*-IN (integrase) and *env* (C2V3) regions of HIV-1 are useful and offer a good landmark for understanding the recombinant genome structure of various HIV-1 subtypes. In our previous reports [35], we showed that the phylogenetic tree based on the *pol* integrase region was very similar to that of the whole *pol* gene. Because of its conserved nature, this portion of the gene was also successfully used to compare evolution among lentiviruses, including simian immunodeficiency virus [7, 16, 35, 49]. As shown by our results, the positions in the trees of the various isolate subtypes that were analyzed consistently correlated with those of the full-length subtype references, and hence may reflect the true evolution of the whole gene structure.

Two dominant HIV-1 subtypes in Cameroon

The most striking and unexpected finding in this study is the two dominant groups of HIV-1 found to be spreading in Cameroon. This appears to be supported by a recent report showing that the neighboring Nigerian strain (IBNG) and two other strains from Djibouti (DJ263 and DJ264) formally classified as subtype A based on their *env* gp120 sequences [20, 21, 30] have a nearly identical A/G mosaic structure [4]. The authors of this report tentatively classified these three strains as "subtype IBNG" because of their possible Nigerian origin. However, we found that the majority of our presumed Cameroonian HIV-1 subtype A isolates tightly clustered in the *env* C2V3 and *pol* integrase analyses together with the above A/G reference strains (Fig. 1 and 2). As shown in the trees, out of 22 Cameroonian isolates classified as subtype A based on *env* sequence, 14 clustered in this group that we are proposing as subtype A2. As suggested by Carr et al., this virus may be more common than the prototypical subtype A in Cameroon and Nigeria [4]. Also, more than half of Cameroonian HIV-1 isolates that were previously classified as subtype A by a partial *env* analysis [42, 54] grouped in this A/G mosaic

CONS-A
94CM323
95CM897
96CG12
97CG276
97CG282
94CM116
94CM142
95CM844
95CM862
95CM884
95CM890
94CM61
94CM90
94CM153
94CM154
94CM157
94CM278
94CM279
95CM371
95CM419
95CM847
96CG13
96CG22
96CG30
96CG38
97CG252
97CG313
97CG315
A
A1
A2
G

type. The close relationship between these subtype A2 isolates and the subtype G cluster in the *pol*-IN phylogeny support the hypothesis that this virus group effectively has a recombinant genome formed from HIV-1 subtypes A and G (Fig. 2). It is also interesting that these viruses have relatively shorter branch lengths in the tree. This might reflect the high speed with which they are spreading in Cameroon. The dynamic role played by the resulting A/G recombination event to form a complex patchwork of related sequences remains unclear. At the *pol* integrase nucleotide sequence level, several common nucleotide residues were found to be shared by subtype G and the new subtype A2 isolates (Fig. 4). Among the common amino acids found in their C2V3 alignment (Fig. 3), the strict conservation of the hexameric sequence GPGQTF was relevant. This particular conserved motif at the apex of the V3 loop sequence strongly suggests that all these mosaic A/G viruses evolved from a common ancestral sequence of *env* subtype A strain. The biological and immunological properties of the mosaic A/G genome viruses, which controls their spread, are unknown. It has been demonstrated that the prinicipal neutralizing determinant (PND) of human immunodeficiency virus resides within the V3 loop of the envelope protein where antibodies elicited by peptides of this region were able to neutralize diverse isolates [22, 25, 38, 47]. It is evident that the conserved hexameric pattern GPGQTF might have some influence on the biological properties of these subtype A2 viruses [59–61]. It has also been demonstrated that sera from two rabbits immunized with a synthetic hexameric peptide GPGRAF neutralized the strains IIIB and MN even though the sequences flanking the GRGRAF region in these isolates are different [22].

Whereas the genomic sequences of this large group of viruses in Cameroon are due to recombination between subtypes A and G, the origin of the second group, which included 6 out of the 22 *env*-related subtype A isolates remains unknown. The *env* sequences of these viruses appeared to be related in the *env* C2V3 tree to prototypical subtype A strain, but their *pol* sequences seemed to have evolved from a completely different virus (Figs. 1 and 2). Detailed analyses of the *pol* integrase nucleotide substitutions excluded the hypothesis of G to A hypermutation, which is known to occur among retroviruses [6, 17, 56] as another possible explanation of their unique position in the *pol*-IN phylogeny (Fig. 4). Rather, we found some residues in this region that are conserved in HIV-1 subtype B, D and even in HIV-1 group O [40] These new viruses were also characterized by their short branch lengths in the *pol* phylogeny (Fig. 2), which is a sign that they were also spreading quickly in the country. Interestingly, two of the Cameroonian isolates (95CM862 and 95CM864) suggest that inter-subtype recombination occurred between the two dominant Cameroonian types of virus (Table 1). The full-length sequences of the isolates of this second group will be needed to clarify whether these viruses

Fig. 4. The *pol* (integrase) nucleotide relationship between the Cameroonian subtype A, A1, A2 and G. The alignment was obtained with the CLUSTRAL W program [58] and minimal manual editing. Dashes indicate sequence identity with the consensus subtype A as recently published by Los Alamos National Laboratory [42]. Several positions of common substitution residues are indicated at the top of the alignment

have evolved from an ancestor virus that is a remombinant between subtype A and another subtypes such as I, J or yet unidentified one. For now, to be cautious, we suggest that this group be called HIV-1 subtype A1 in order to show that its *env* sequence is related to that of subtype A strains.

Other HIV-1 group M subtypes

In the phylogenetic trees, the 49 Central African HIV-1 isolates fell into distinct lineages suggesting that four of them (HIV-1 subtypes D, F, G and H) evolved as non-recombinant viruses. In the *pol*-IN tree, the cluster formed by the subtype B references was distinct, even though it was completely included in the D cluster. In this region, the subtype B remained tightly clustered within the B/D group (bootstrap value 95%) (Fig. 2). This reflect the close relationship between HIV-1 subtypes B and D. It has been demonstrated that for most parts of the HIV-1 genome, subtypes B and D are more closely associated with each other than any other subtypes [31, 49, 62]. Our 5 subtype D isolates (two Cameroonian isolates and three Congolese isolates) displayed properties of older diverged sequences as shown in the trees (Fig. 1 and 2). Their viral sequences in the *env* C2V3 region were very divergent and branched off before the former Zairian subtype D reference (NDK) that has been considered as one of the divergent D strains [53, 62]. Furthermore our *pol*-IN analysis confirmed the divergence of these five isolates (Fig. 2). Subtype D has been reported to be more prevalent in East Africa and the nighboring former Zaire [44, 45, 52]. It is possible that the original ancestor of the current epidemic of HIV-1 subtype D found in this region of Africa is not far from our present D strians.

Also interesting are the results of the subtype G cluster found in our *pol*-IN analysis. Many speculations have been made about the rarity of the prototypical subtype G virus [4, 15, 40]. Recently, two independent research groups reported the sequences of three full-length genomes that they described as possible non-recombinant subtype G isolates. Remarkably, all 9 of our HIV-1 isolates of *env* subtype G formed a very distinct cluster in the *pol*-IN tree together with the subtype G references. Although it is true that a substantial G sequence has recombined with subtype A in Africa [1, 3, 10, 13], the present 9 isolates, which clustered in the trees together with possible G references, may indicate that subtype G has continued to spread as an integral structure through Central Africa at a considerable rate. Consequently, the parental G viruses probably existed and might not be as rare as previously thought [4, 15].

Subtype H is one of the minority subtypes within the group M of HIV-1. Information on the *pol*-IN sequence of subtype H was limited to only one strain [40]. We report here 7 new subtypes H (2 from Cameroon and 5 from Congo). These isolates clustered in the *pol*-IN and *env* analyses (Fig. 1 and 2) together with the recent and unique full-length sequence reference subtype H, strain 90CF056 from Central Africa Republic [15]. Furthermore, the *pol*-IN subtype H cluster identified here also correlated well with the results of the *env* tree strongly suggesting that the genome structure of these subtype H isolates evolved independently without

any recombination with another HIV-1 subtype. The C2V3 amino-acid alignment of these H isolates showed several conserved residues that will be useful for future redefinition of the subtype H consensus (Fig. 3).

Acknowledgement

This work was supported by Grants-in-Aid from the Japanese Ministry of Education, Science, Sports and Culture.

References

1. Abimiku AG, Stern TL, Zwandor A, Markham PD, Calef C, Kyari S, Saxinger WC, Gallo RC, Robert-Guroff M, Reitz MS (1994) Subgroup G HIV type 1 isolates from Nigeria. AIDS Res Hum Retroviruses 10: 1 581–1 583
2. Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlen M (1994) Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. J Virol 68: 5 918–5 924
3. Bobkov A, Cheingsong-Popov R, Karasyova N, Selimova L, Kravchenko A, Bochkova M, Kazennova E, Ladnaya N, Pokrovsky V, Weber J (1996) Sequence analysis of the Glycoprotein 120 coding region of a new human immunodeficiency virus type 1 subtype G strain from Russia. AIDS Res Hum Retroviruses 12: 1 385–1 388
4. Carr JK, Salminen MO, Albert J, Sanders-Buell E, Gotte D, Birx DL, McCutchan FE (1998) Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 247: 22–31
5. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, St. Louis D, Burke DS, McCutchan FE (1996) Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J Virol 70: 5 935–5 943
6. Cheynier R, Gratton S, Vartanian J-P, Meyerhans A, Wain-Hobson S (1997) G to A hypermutaion does not results from polymerase chain reaction. AIDS Res Hum Retroviruses 13: 985–986
7. Clewley JP, Lewis JCM. Brown DWG, Gadsby EL (1998) A novel simian immunodeficiency virus (SIV drl) *pol* sequence from the Drill monkey, Mandrillus leucophaeus. J Virol 72: 1 0305–1 0309
8. Coffin JM (1995) HIV population dynamics in vivo: Implications for genetic variation, pathogenesis and therapy. Science 267: 483–489
9. Cornelissen M, van den Burg R, Zorgdrager F, Lukashav V, Goudsmit J (1997) *pol* gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtype B and D. J Virol 71: 6 348–6 358
10. Duchet S, Letourneur F, Loussert-Ajaka I, Chaplain C, Gomes E, Brunvezinet F, Simon F, Saragosti E (1997) *Gag* and *env* sequence of an A/G/H Recombinant from a Zairian HIV type 1 isolate. AIDS Res Hum Retroviruses 13: 1 351–1 354
11. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791
12. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH (1999) Origin of HIV-1 in the chimpanzee pan troglodytes troglodytes. Nature 397: 436–441
13. Gao F, Robertson DL, Carruthers CD, Li Y, Bailes E, Kostrikis LG, Salminen MO, Bibollet-Ruche F, Peeters M, Ho DD, Shaw GM, Sharp PM, Hahn BH (1998) An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a

complex mosaic comprising three different group M subtype (A, G, and I). J Virol 72: 1 0234–1 0241
14. Gao F, Robertson DL, Morrison SG, Hui H, Craig S, Decker J, Fultz PN, Girard M, Shaw GM, Hahn BH, Sharp PM (1996) The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J Virol 70: 7 013–7 029
15. Gao F, Robertson DL, Carruthers CD, Morrison SG, Jian B, Chen Y, Barre SF, Girard M, Srinivasan A, Abimiku AG, Shaw GM, Sharp PM, Hahn BH (1998) A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J Virol 72: 5 680–5 698
16. Gojobori T, Moriyama EN, Ina Y, Ikeo K, Miura T, Tsujimoto H, Hayami M, Yokoyama S (1990) Evolutionary origin of human and simian immunodeficiency viruses. Proc Natl Acad Sci USA 87: 4 108–4 111
17. Gunther S, Sommer G, Plikat U, Iwanska A, Wain-Hobson S, Will H, Meyerhans A (1997) Naturally occurring Hepatitis B virus genomes bearing the hallmarks of retroviral G to A hypermutation. Virology 235: 104–108
18. Gurtler L, Hauser P, Eberle J, Von Brunn A, Knapp S, Zekeng L, Tsague J, Kaptue L (1994) A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol 68: 1 581–1 585
19. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373: 123–126
20. Howard TM, Olaylele DO, Rasheed S (1994) Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria. AIDS Res Hum Retroviruses 10: 1 755–1 757
21. Howard TM, Rasheed S (1996) Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strains from Nigeria. AIDS Res Hum Retroviruses 12: 1 413–1 425
22. Jayaherian K, Langlois AJ, LaRosa GJ, Profy AT, Bolognesi DP, Herlihy WC, Putney SD, Matthews TJ (1990) Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. Science 250: 1 590–1 593
23. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120
24. Kostrikis L, Bagdades E, Cao Y, Zhang L, Dimitriou D, Ho D (1995) Genetic analysis of Human Immunodeficiency virus type 1 strains from patients in cyprus: identification of a new subtype designated subtype 1. J Virol 69: 6 122–6 130
25. Larosa GJ, Davide JP, Weinhold K, Waterbury JA, Profy AT, Lewis JA, Langlois AJ, Dreesman GR, Boswell RN, Shadoluck P, Holley LH, Karplus M, Bolognesi DP, Matthews TJ, Emini EA, Putney SD (1990) Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science 249: 932–935
26. Leitner T, Alaeus A, Marquins S, Lilja E, Lidmann K, Albert J (1995) Yet another subtype of HIV type 1? AIDS Res Hum Retroviruses 11: 995–997
27. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J (1996) Accurate reconstruction of known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci USA 93: 1 0864–1 0869
28. Leitner T, Escanilla D, Marquina S, Wahlberg J, Brostrom C, Hansson H, Uhlen M, Albert J (1995) Biological and molecular characterization of subtype D, G and A/D recombination HIV-1 transmissions in Sweden. Virology 209: 136–146

29. Loussert-Ajaka I, Chaix M-L, Korber B, Letourneur F, Gomas E, Allens E, Ly T-D, Brun-Vezinet F, Simon F, Saragosti S (1995) Variability of human immunodeficiency virus type 1 group O strains ioslated from Cameroonian patients living in France. J Virol 69: 5 640–5 649
30. Louwagie J, Janssens W, Mascola J, Heyndrickx L, Hegerich P, Van Der Groen G, McCutchan FE, Burke DS (1995) Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J Virol 69: 263–271
31. Louwagie J, McCutchan FE, Peeters M, Brennan TP, Sanders-Buell E, Eddy GA, van der Groen G, Fransen K, Gershy-Damet G-M, Deleys R, Burke DS (1993) Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS 7: 769–780
32. Mahieux R, Ibrahim F, Mauclaire P, Herve V, Michel P, Tekaia F, Chappey C, Garin B, Van Der Ryst E, Guillemain B, Ledru E, Delaporte E, De The G, Gessain A (1997) Molecular epidemiology of 58 new African human T-cell leukemia virus type 1 (HTLV-1) strains: Identification of a new and distinct HTLV-1 molecular subtype in Central Africa and in Pygmies. J Virol 71: 1 317–1 333
33. Mboudjeka I, Zekeng L, Yamashita M, Takehisa J, Ido E, Miura T, Ohkura S, Ikeda M, Kaptue L, Hayami M (1997) Prevalence and phylogenetic analysis of HTLV-1 isolates in Cemeroon, including those of the Baka Pygmy. Jpn J Cancer Res 88: 619–624
34. McCutchan FE, Salminen MO, Carr JK, Burke DS (1996) HIV-1 genetic diversity. AIDS 10 [Suppl 3]: s13–s20
35. Miura T, Sakuragi J, Kawamura M, Fukasawa M, Moriyama EN, Gojobori T, Ishikawa K, Mingle JAA, Nettey VB, Akari H, Enami M, Tsujimoto H, Hayami M (1990) Establishment of a phylogenetic survey system for AIDS-related lentiviruses and demonstration of a new HIV-2 subgroup. AIDS 4: 1 257–1 261
36. Mohri H, Singh HM, Ching WT, Ho DD (1993) Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the blood of treated and untreated patients. Proc Natl Acad Sci USA 90: 25–29
37. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo H-M, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman R, Ho DD, Boucher CAB, Leonard JM, Norbeck DW, Kempf DJ (1996) Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. Nature Med 2: 760–766
38. Moore J, Trkola A (1997) HIV type 1 coreceptors, neutralization serotypes, and vaccine development. AIDS Res Hum Retroviruses 13: 733–736
39. Myers G (1994) HIV: Between past and future. AIDS Res Hum Retroviruses 10: 1 317–1 324
40. Myers G, Korber B, Foley B, Jeang K-T, Mellors JW, Wain-Hobson S (1996) Human Retroviruses and AIDS: a compilation and analysis of nuclei acid and amino acid sequences. Theoretical Biology and Biophysics group, Los Alamos National Laboratory, Los Alamos
41. Najera I, Holguin A, Quinones-Mateu ME, Munoz-Ferandez MA, Najera R, Lopez-zalindez, Domingo E (1995) *pol* gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. J Virol 69: 23–31
42. Nkengasong JN, Janssens W, Heyndrickx L, Fransen K, Ndumbe PM, Motte J, Leonaers A, Ngolle M, Ayuk J, Piot P, van der Groen G (1994) Genotypic subtypes of HIV-1 Cameroon. AIDS 8: 1 405–1 412
43. Page RD (1996) An application to display to phylogenetic trees on personal computers. Comput Appl Biosci 22: 357–358

44. Ranjbar S, Slade A, Jenkins A, Heath A, Kitchen P, Almond N, Osmanov S, Holmes D (1995) Molecular characterization of an HIV type 1 isolate from Burundi. AIDS Res Hum Retroviruses 11: 981–984
45. Rayfield MA, Downing RG, Baggs J, HU DJ, Pieniazek D, Luo C-C, Biryahwaho B, Otten RA, Sempala SD, Dondero TJ, The HIV Variant Working Group (1998) A molecular epidemiologic survey of HIV in Uganda. AIDS 12: 521–527
46. Robertson DL, Sharp PM, McCutchan FE, Hahn BH (1995) Recombination in HIV-1. Nature 374: 124–126
47. Rusche JR, Javaherian K, Mcdanal C, Petro J, Lynn DL, Grimaila R, Langlois A, Gallo RC, Arthur LO, Fischinger PJ, Bolognesi DP, Putney SD, Matthews TJ (1988) Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. Proc Natl Acad Sci USA 85: 3 198–3 202
48. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
49. Sharp P, Robertson D, Gao F, Hahn BH (1994) Origins and diversity of human immunodeficiency virus. AIDS 8: s27–s42
50. Simmonds P, Zhang LQ, McOmish F, Balfe P, Ludlam CA, Brown JAL (1991) Discontinuous sequence change in human immunodeficiency virus (HIV) type 1 *env* gene sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J Virol 65: 6 266–6 276
51. Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barre-Sinoussi F, Vezinet B (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nature Med 4: 1 032–1 037
52. Siwka W, Schwinn A, Baczko K, Pardowitz I, Mhalu F, Shao J, Rethwilm A, Meulen VT (1994) *Vpu* and *env* sequence variability of HIV-1 isolate from Tanzania. AIDS Res Hum Retroviruses 10: 1 753–1 754
53. Spire B, Sire J, Zachar V, Barre-Sinoussi F, Galibert F, Hampe A, Cherman JC (1989) Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. Gene 81: 275–284
54. Takehisa J, Zekeng L, Ido E, Mboudjeka I, Moriyama H, Miura T, Yamashita M, Gurtler LG, Hayami M, Kaptue L (1998) Various type of HIV mixed infections in Cameroon. Virology 245: 1–10
55. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4 673–4 680
56. Vartanian J-P, Meyerhans A, Asjo B, Wain-Hobson S (1991) Selection, recombination, and G-A hypermutation of human immunodeficiency virus type 1 genomes. J Virol 65: 1 779–1 788
57. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, Saag MS, Shaw GM (1995) Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373: 117–122
58. WHO Network for HIV Isolation and Characterization (1994) HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. AIDS Res Hum Retroviruses 10: 1 327–1 343
59. Willey RL, Smith DH, Laskey LA, Theodore TS, Earl PL, Moss B, Capon DJ, Martin MA (1998) In vitro mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. J Virol 62: 139–147

60. Wolfs TFW, Zwart G, Bakker M, Valk M, Kuiken CL, Goudsmit J (1991) Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology 185: 195–205
61. Zhang LQ, MacKenezie P, Cleland A, Holmes EC, Leigh Brown AJ, Simmonds P (1993) Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J Virol 67: 3 345–3 356
62. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391: 594–597

Authors' address: Dr. T. Miura, Laboratory of Viral Pathogenesis, Institure for Virus Research Kyoto University, 53 Shogoin-Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan.

Received May 31, 1999