Complete nucleotide sequence of an African human T-lymphotropic virus type II subtype b isolate (HTLV-II-Gab): molecular and phylogenetic analysis

Franck Letourneur,1 Luc d’Auriol,1 Marie-Christine Dazza,2 Martine Peteers,3 Issa Bedjabaga,4 Peter Piot,3 Eric Delaporte,2 Antoine Gessain5 and Nicole Monplaisir1

1 Genset, 1 passage E. Delaunay, 75011 Paris, France
2 INSERM U13/Fondation Léon Mba, Paris, France
3 Institute of Tropical Medicine, Antwerp, Belgium
4 Centre International de Recherches Médicales, Franceville, Gabon
5 Unité d’Épidémiologie des Virus Oncogènes, Institut Pasteur, Paris, France

We report the first complete nucleotide sequence of an African human T-cell lymphotropic virus type II. This new strain, called HTLV-II-Gab (Gab), was obtained from the uncultured peripheral blood mononuclear cells of a 44-year-old healthy Gabonese male who lived in a remote rural area, with neither history of blood transfusion nor sexual intercourse with non-Africans. Using nested PCR, 25 overlapping fragments, representing the entire proviral genome, were obtained, cloned and sequenced. The overall nucleotide sequence comparison with the four other available complete HTLV-II genomes indicated that Gab was more closely related to the HTLV-II subtype b prototypes (98.9, 99.3 and 98.2% nucleotide similarity with G12, NRA and GU respectively) than to the subtype a prototype (95.1% nucleotide similarity with Mo). Restriction profiles studies and phylogenetic analyses confirmed that Gab was a subtype b strain. However, this strain represents a newly described restriction fragment length polymorphism subtype, closely related to one of the rare partially sequenced African isolates originating from a pygmy living in Cameroon (PYGCAM). Nevertheless, the very low genetic divergence observed between this new African strain and the American strains raises several questions on the origins and level of genetic variability over time of this human retrovirus.

Introduction

The human T-cell lymphotropic viruses, type I (HTLV-I; Poiesz et al., 1980) and type II (HTLV-II; Kalyanaraman et al., 1982), are closely related members of a group of mammalian retroviruses sharing common epidemiological, virological and molecular characteristics. HTLV-I infection is mainly associated with adult T-cell leukaemia (ATL; Poiesz et al., 1980), a malignancy of mature activated CD4 lymphocytes, and with a chronic neurological disorder, known as tropical spastic paraparesis (TSP)/HTLV-I associated myelopathy (HAM) (Gessain et al., 1985). In contrast, despite recent studies which have linked HTLV-II infection with a spectrum of neurological and possibly lymphoproliferative disorders, there is still no clear evidence that HTLV-II causes any human disease (Fouchard et al., 1995; Hall et al., 1996).

Numerous sero-epidemiological studies have shown that HTLV-II is highly endemic in many New World indigenous populations, but also epidemic-endemic among intravenous drug users (IVDU) from the United States and, to a lesser extent, from South European countries (Hall et al., 1996; Salemi et al., 1995, 1996). Furthermore, since 1991, sporadic cases of HTLV-II infection have been detected in West and Central Africa, and the presence of such infection in rural isolated populations, including pygmies, suggests an ancient presence of HTLV-II in this area (Gessain & de The, 1996).

Study of the genetic heterogeneity of HTLV-II was first based on the nucleotide sequence divergence and restriction

Author for correspondence: Franck Letourneur. Present address: INSERM IFR18 ICGM, 22 rue Mechain, 75014 Paris, France. Fax +33 1 40 51 64 73. e-mail franck@cochin.inserm.fr

The nucleotide sequence reported in this paper has been submitted to the EMBL data library and assigned accession number Y13051.
mapping of the envelope transmembrane protein gp21. This suggested that there were at least two closely related but genetically distinct subtypes of HTLV-II, designated a and b (Hall et al., 1992; Pardi et al., 1993 b). This env-based clustering was confirmed by phylogenetic studies on a fragment of the reverse transcriptase pol gene (Dube et al., 1994). However, the gp21 region of HTLV-II shows such high sequence homology that discrimination within the two viral subgroups was impossible (Hjelle et al., 1993). In contrast, analysis of the most divergent proviral region – the long terminal repeats (LTR) (Takahashi et al., 1993) – by restriction fragment length polymorphism (RFLP) methods and phylogenetic analysis has permitted a more precise discrimination of the genetic diversity in HTLV-II. The good correlation between the two methods led to the recognition within both subtypes of different phylogroups, each carrying one or more distinct restriction profiles. These different phylogroups were predominantly resolved on the basis of geography and/or ethnicity (Eiraku et al., 1995; Switzer et al., 1995 b).

Currently, the epidemiological repartition of the two subtypes appears less simple than previously thought: indeed, both subtypes are present in the Amerindian tribes of Central and South America (Pardi et al., 1993 b; Eiraku et al., 1998; Hall et al., 1996; Heneine, 1996) and in the North American Indian groups (Hjelle et al., 1993). Subtype a is clearly the predominant subtype infecting IVDU in urban areas of North America (Hall et al., 1992), and subtype b is mainly present in IVDU from South European countries (Hall et al., 1996; Salemi et al., 1995, 1996).

Only very few molecular data are available from African HTLV-II.

1. Two subtype a isolates have been identified in human immunodeficiency virus type 1 (HIV-1) co-infected prostitutes from Ghana (GhK; Igarashi et al., 1993) and Cameroon (PH230PCAM; Maucle et al., 1995).

2. Three HTLV-II subtype b isolates have been partially characterized: the first originated from a plasma of a Zairian patient collected in 1969 (Dube et al., 1994). The second originated from a Cameroonien pygmy of the Bakola tribe (PYGCAM; Gessain et al., 1995) and was shown to be very close to the American G12 isolate. The third originated from several members of a Gabonese family (JPS; Tuppin et al., 1996) and represented the most divergent HTLV-IIb isolate yet described.

We report here the fourth complete nucleotide sequence of an HTLV-II isolate and the first complete one of African origin and discuss the phylogenetic relationships between our new isolate and all the available HTLV-II sequences, especially those of African origin.

Methods

**Origin of the patient.** In 1989, a sero-epidemiological study was carried out in Libreville (Gabon) on 322 adults screened for HTLV-I and -II infection (Delaporte et al., 1991) after informed consent and medical examination. Among the 25 Western blot indeterminate samples, a simultaneous PCR study on both the gag and pol regions identified one HTLV-II positive sample. This belonged to a 44-year-old man, living in a remote rural area around Libreville. He was asymptomatic with neither history of blood transfusion, parenteral drug use nor sexual intercourse with non-Africans. No information was available for the rest of his family.

**DNA preparation.** Peripheral blood mononuclear cells (PBMC) were separated by centrifugation on a Ficoll–Hypaque gradient and frozen in 10% DMSO in liquid nitrogen. DNA was isolated from cells, in a laboratory free of any HTLV-II DNA, by lysis in a buffer containing 10 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.5% SDS, 2 mM EDTA and 100 µg/ml protease K. The DNA preparation was incubated overnight at 37 °C, followed by two extractions with phenol–chloroform–isoamyl alcohol (25:24:1) and precipitation with ethanol. The DNA pellet was vacuum dried and redissolved in water.

**PCR amplifications.** Nested PCR amplifications were performed in 100 µl reaction mixtures containing 200 µM each deoxynucleoside triphosphate (Pharmacia), 10 mM Tris–HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 2.5 U Taq polymerase (Perkin Elmer Cetus) and 20 pmol of each primer (Genset). Each initial reaction contained 1 µg of DNA and 5 µl of the first round PCR product was used in the second round. PCR primer sequences were obtained using the HTLV-II Mo complete nucleotide sequence (Shimotohno et al., 1985; GenBank no. M10060) as reference with Oligo software 4.0. The reactions were carried out in a DNA thermocycler (Perkin Elmer) for 40 cycles. The denaturation, annealing and elongation conditions were 94 °C (for 30 s), between 50 and 60 °C according to the primers used (for 30 s) and 72 °C (for 1 min) respectively. We thus generated 25 overlapping fragments (from 101 to 702 bp) spanning the entire genome. The sequences of the primers were as follows (sequence positions are according to the Mo complete nucleotide sequence): (1) 201 (bp 28–47), AGCCACCCAGGCGGAGTCAT, and 202 (bp 468–487), CTAAAGGGCGCCGCAGCTTC; (2) 205 (bp 908–927), GCAGCCCTAGGCGCCTTGGATT, and 206 (bp 1521–1540), AGACCTTCTGGGGCGGGGTT; (3) 203 (bp 417–436), GGCCTCGCTGCTGGGCGGGTG; (4) 207 (bp 1491–1508), ATGCCAGGCCCCTAAAAAG, and 208 (bp 1595–1785), TCTTGGCAGATTGGGCCCCAGG; (5) 209 (bp 1879–1898), CCCCCACACAGCCTTCTT, and 210N (bp 2156–2175), GTGTTGCTTGGTCCTCCGCG; (6) 20B (bp 2040–2060), TCTGTTGTCAGTCTTCTGCC, and 210BIS (bp 2439–2459), GGGCGCTTCTCCAAATGGT; (7) 239 (bp 2381–2400), TTCCGGCTTCTCCAAATGGT; (8; 19a) (bp 2973–2992), CCCCCACACAGCCTTCTT, and 212 (bp 3464–3483), AGTGGAGGCCGCGGTTGAGCGG; (9) 22.1 (bp 3705–3721), CCTTTTGGACTTCTGGA, and 214 (bp 4053–4072), TAAAGGAGGCACTTGGTG; (10) OLIGO1 (bp 5995–5913), AAAAAACATCTGGCCAAAAAA and 31.2 (bp 4342–4361), GGCCTCATACGCTAGCTGAG; (11) 217 (bp 4522–4541), TCGCCGCCGCTGGTCGCTGGA, and 219R (bp 5130–5149), CAGATTCTGCCTGGCTTCTGAC; (12) 31.1 (bp 4282–4301), GCCTCCCTGGTCTGGCGCTTC; (13) 219 (bp 4989–5008), TAAACTCCCCCGCTTACCA, and 218BIS (bp 5228–5252), TCGTGATCTGCTCTGTTGG; (14) 25.1BIS (bp 5253–5272), TCAGATTGTATCTCTCCTCC, and 25.1BISREV (bp 5409–5428), ATGTTGGAAATGACTTATAA; (15) 25.1 (bp 5232–5252), CCCACAGGGCTGACAGCAG, and 220BIS (bp 5478–5497), TAGCGAGGAAAGCTTGTGAG; (16) 26.5 (bp 5469–5489), GGCTCCCTTACAATGACATCC, and 222RE (bp 5993–6012), TGTAGGCGAGTTGGGAGA; (17) FL34S (bp 6284–6303), CAGTATGACCCGAAAAATAG and 223REV (bp 6696–6715), GTTATAGAGGACTTGATGAT.
African HTLV-II subtype b isolate

AAACAAACCTCCCTCCGAACC and 224BIS (bp 6476–6492), GTCTCCCTCAGGCTATATCC (19) 226REV (bp 6774–6793), TCCCTCAACCGGTGTCCTCAAC and 226 (bp 6906–6927), GTTGGGACGGCGTGTAAG and 227BIS (bp 6820–6839), TTAATGGATTCTCCTGGAGAGG; (21) 227 (bp 6907–6926), CTITCCCTCCGGCGGTGTT and 227REV (bp 7247–7266), CAACAGGCGGATCATCAT; (22) 228REV (bp 7144–7163), GGTCTCCTAACGGCAATCTC and 228 (bp 7616–7635), GTTGGGACGGCGTGTAAG; (23) 229BIS (bp 7591–7610), CTTCCCTCCGGCGGTGGAAAGCT; (24) 229BIS (bp 7591–7610), CTTCCTCCGCCGCCCTCC and 229BIS (bp 7918–7937), GTTGGGACGGCGTGTAAG; (20) 235 (bp 6704–6723), GTCCTCTATACCAGATGAGT and 235 (bp 6704–6723), GTCCTCTATACCAGATGAGT; (19) 226REV (bp 6774–6793), TTCCTCTAACGGAATCTC and 226REV (bp 6774–6793), TTCCTCTAACGGAATCTC.

Molecular cloning and sequencing. PCR fragments were purified from low melting point agarose gel, kinased with T4 polynucleotide kinase and ligated to Smal-digested and dephosphorylated M13mp18 DNA. After transformation of E. coli strain TG1, DNA was extracted from white plaques on IPTG–X-Gal plates. Sequences were obtained using an ABI 373A DNA sequencer (Applied Biosystems) with dye terminator chemistry. Some PCR products were directly sequenced after purification. The M13 universal primer and 40 specific primers were used so that each base was sequenced at least once in both directions.

Nucleotide sequence alignment and analysis. Alignment was performed using Geneworks software (Intelligenetics). The sequence obtained was compared to nucleotide and amino acid sequences of the prototypic HTLV-II subtype b isolates NRA (Lee et al., 1993) and G12 (Pardi et al., 1993 b), the recently described first complete European isolate Gu (Salemi et al., 1996) and the prototypic subtype a isolate Mo (Shimotohno et al., 1985). Mo and NRA originated from two different North American patients, both with hairy cell leukaemia, Gu originated from an Italian IVDU and G12 from a healthy Guaymi Indian in Panama. The nucleotide sequence analysis was performed with Geneworks and DNA Strider 1.2 software. The nucleotide positions numbered in the text are based on the Mola nucleotide sequence.

Phylogenetic analysis. We restricted our phylogenetic analyses of the env gene to gp21 because of the larger number of known gp21 fragments (approximately 30), relative to the low number of complete env sequences (only five). Nearly all the available HTLV-II env gp21 and LTR sequences were used. Concerning the env gp21 tree, the different isolates used were DOG, GAR, PAR (Hall et al., 1992), 130P, Msa1B, 408N, 72969N (Hjelle et al., 1993), Bo, Md, Va (Salemi et al., 1995), JPS (Tupin et al., 1996), Kay1 and 2 (Ishak et al., 1995), PH230PCAM (Maurel et al., 1995), SP1 and 2 (Eiraku et al., 1996).

The LTR sequences originated from isolates ATL18, BRAZ.A21, LA8A, NAV.DS, NOR2N, PUEB.AG and PUEB.BB, ITA47A and 50A, PEN7A, SEM1050 and 1051, SPAN129 and 130 (Switzer et al., 1995 b), WYU1 and 2 (Switzer et al., 1995 a), ED and JG (Takahashi et al., 1993), Va (Salemi et al., 1995), PH230PCAM (Maurel et al., 1995), PYGCAM1 (Gessain et al., 1995), GHKt (Igarashi et al., 1993), MEK17, KAY73 and 139 (Switzer et al., 1996).

Sequence alignment was done with ClustalW (Higgins et al., 1992). To obtain a valid phylogenetic reconstruction, both trees were constructed by two different methods: (i) the neighbour-joining method (NJ) of Saitou & Nei (1987) contained in the ClustalW package (Higgins et al., 1992) – the reliability of the tree was statistically evaluated by using 1000 bootstrap replicates (Felsenstein, 1985); (ii) the maximum likelihood method (ML) using the FastDNAML program, which utilizes randomized data input and global rearrangement. The Treeeop program was used to display the phylogenetic tree.

We used the HTLV-I ATK strain (Seiki et al., 1983) as an outgroup for phylogenetic analysis on both the envelope transmembrane protein gp21 (env gp21) and the LTR.

The two regions studied were the complete env gp21 and the LTR sequences corresponding to nucleotide 86–700.

Results

Generation of overlapping PCR fragments

Genomic DNA from uncultured PBMC was amplified to obtain 25 overlapping PCR fragments, representing the entire genome. A schematic representation of the strategy used is depicted in Fig. 1.

Overall genetic organization and variability of HTLV-II Gab

The entire HTLV-II Gab nucleotide sequence showed closer nucleotide identity with the subtype b isolates (98±9 and 98±2 % with G12 and Gu), and particularly NRA (99±3 %), than with Mo, the prototype of subtype a (95±1 %) (Table 1).

As already observed (Takahashi et al., 1993), the greatest divergence between the different isolates occurred in the LTR (Table 1). We found 53 nucleotide differences between Gab and Mo (6±9 % nucleotide divergence) and from 9 to 13 nucleotide differences between Gab and the representative IIb (1±2–1±8 % nucleotide identity).

The important regulatory elements of the U3 region (21 bp repeats, polyadenylation signal, TATA box, mRNA cap site) were well-conserved. As observed for the other HTLV-IIb isolates, a C → A mutation on the second base of the second repeat (nucleotide position 133) resulted in the formation of a DraII restriction site specific for subtype b (Switzer et al., 1995 b). The highest nucleotide sequence conservation was observed in the R region, preserving the Rex responsive element (RRE). The primer binding site was also maintained.

Comparison of the gag nucleotide sequence showed greater identity between Gab and subtype b isolates (99±31, 98±92 and 97±77 % nucleotide identity with NRA, G12 and Gu). The close proximity of Gab with NRA was confirmed by the amino acid comparison (Table 1).

The nucleotide sequence of the pol gene of Gab showed 4±8 % variation (49 non-synonymous substitutions (NSS)) from isolate Mo, 0±6 % from NRA (5 NSS), 0±75 % from G12 (8 NSS) and 2±24 % from Gu (28 NSS). We identified the same C → T substitution at position 97 (from the start of the pol gene) already observed with NRA, G12 and Gu (Salemi et al., 1996), thus resulting in the insertion of a stop codon at amino acid position 32.

Comparison of the env gene nucleotide sequence of Gab with that of the other isolates showed diversity ranging between 95±48 % for Mo and 98±97 % for G12 and NRA. The amino acid sequence divergence based on the env gene sequences available from the two other African HTLV-IIb isolates known to date with the two prototypic subtype b
Table 1. Overall and detailed nucleotide and amino acid sequence comparison of the new HTLV-II-Gab with the three other known HTLV-II complete sequences

(a) Nucleotide sequence similarity (%)

<table>
<thead>
<tr>
<th>HTLV-II-Gab compared with:</th>
<th>Overall</th>
<th>LTR</th>
<th>gag</th>
<th>pol</th>
<th>env</th>
<th>pX</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTLV-II-Mo</td>
<td>95:1</td>
<td>93:08</td>
<td>95:7</td>
<td>95:12</td>
<td>95:48</td>
<td>95:40</td>
</tr>
<tr>
<td>HTLV-II-Gu</td>
<td>98:2</td>
<td>98:17</td>
<td>97:77</td>
<td>97:77</td>
<td>98:9</td>
<td>98:26</td>
</tr>
</tbody>
</table>

(b) Amino acid sequence similarity (%)

<table>
<thead>
<tr>
<th>HTLV-II-Gab compared with:</th>
<th>gag</th>
<th>pol</th>
<th>env</th>
<th>tax</th>
<th>rex</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTLV-II-Mo</td>
<td>98:38</td>
<td>95:01</td>
<td>97:73</td>
<td>97:47</td>
<td>94:12</td>
</tr>
<tr>
<td>HTLV-II-NRA</td>
<td>100</td>
<td>99:49</td>
<td>98:55</td>
<td>99:16</td>
<td>99:41</td>
</tr>
<tr>
<td>HTLV-II-Gu</td>
<td>98:01</td>
<td>97:15</td>
<td>98:77</td>
<td>97:19</td>
<td>97:06</td>
</tr>
</tbody>
</table>

isolates NRA and G12 and the recently described Gu is given in Table 2. Gab is closer to the Amerindian G12, to the Italian IVDU Gu and to the Cameroonian pgynm PYGCAM (1:23, 1:23 and 1:44% amino acid divergence respectively) (Gessain et al., 1995) than to JPS, the other Gabonese isolate (2:47% amino acid divergence) (Tuppin et al., 1996). No amino acid position encoded by the env sequences was common for all the African HTLV-III strains known to date. However, His-278 and Ser-442 (from the start of the env-encoded amino acid sequence) were specific for Gab. The immunodominant B cell epitope defined by Pardi et al. (1993b) contained the same amino acid substitutions already observed for G12 and Gu (Salemi et al., 1996): Ser-183 → Pro and Ile-206 → Met. The amino acid sequence of the immunodominant T cell epitope K55 (Lipka et al., 1992) was conserved. The cysteine residues (at positions 389, 396 and 397 from the amino terminus of the Env frame) involved in S–S bridges between surface glycoprotein and transmembrane proteins were well-conserved too.

The tax gene comprised 1071 nucleotides encoding 356 amino acids. A change of two nucleotides, TA → CA, towards the 3′ end of the tax gene inserted an arginine residue for a stop codon normally present at position 332 of the Mo Tax protein, resulting in a protein 25 amino acids longer. This extended Tax protein has been shown to be a unique characteristic of HTLV-II subtype b (Pardi et al., 1993a).

Interestingly, as observed for G12, NRA and Gu, the Rex protein had the lowest percentage amino acid sequence identity when compared with the Mo Rex protein (94:12%). The arginine-rich region in the amino terminus and the strongly conserved amino acid sequence 55–70 were unchanged in Gab.

A detailed study of the pX region of the Gab isolate demonstrated the presence of five ORFs, encoding four accessory proteins (Fig. 1), as described initially for the Mo isolate by Ciminale et al. (1992, 1995). These proteins were designated according to their size and coding ORFs: p10^{pX1}, p12^{pX2}, p15^{pX3}, p21^{pX4}, and p16^{pX5}.
Table 2. Amino acid sequence comparison of the env-encoded protein of HTLV-II-Gab with those of other HTLV-II isolates

<table>
<thead>
<tr>
<th>Amino acid divergence (%)</th>
<th>PYGCAM</th>
<th>JPS</th>
<th>G12</th>
<th>Gu</th>
<th>NRA</th>
<th>Mo</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTLV-II-Gab (Gabon)</td>
<td>1.44</td>
<td>2.47</td>
<td>1.23</td>
<td>1.23</td>
<td>1.45</td>
<td>2.27</td>
</tr>
<tr>
<td>PYGCAM (Cameroon)</td>
<td>2.26</td>
<td>1.02</td>
<td>0.62</td>
<td>1.23</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>JPS (Gabon)</td>
<td>1.65</td>
<td>1.65</td>
<td>2.26</td>
<td>3.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enzyme restriction analysis

Previous reports have demonstrated the ability of RFLP analysis to differentiate between the two subtypes: indeed, analysis of the nucleotide sequence of the env gene indicated the presence of a C → T mutation at nucleotide position 1030 (from the start of the env gene) of Gab. This mutation disrupted a Xhol restriction site normally found in the nucleotide sequence of the env-encoded gp21 of Mo and used by some to discriminate the two subtypes (Hall et al., 1992).

Recent studies on the LTR have identified restriction sites specific to each subtype. Indeed, DraII and AvaII sites (at nucleotide positions 133 and 706 respectively) are specific to subtype b. Two different detailed analyses of the LTR, using different restriction sites, have permitted classification of the subtype b strains into six (b0–b5) restriction types according to Switzer et al. (1995b) and five (b1–b5) restriction types according to Eiraku et al. (1995). Nucleotide sequence analysis of the LTR of Gab revealed the presence of the DraII and AvaII sites specific for subtype b. We observed for Gab a restriction profile identical to b3 of Eiraku and close to, but slightly different from, b5 of Switzer (difference due to the presence of a new BanII site at nucleotide position 385) (Fig. 2).

Phylogenetic analysis of HTLV-II Gab

In the phylogenetic trees constructed by the NJ method on the 588 bp of the env gp21 region, the two subtypes a and b were resolved with high bootstrap values (99 and 90% respectively) and Gab belonged to subtype b, close to the
sequences argued for an absence of recombination between these two regions.

**Discussion**

The first complete nucleotide sequence of an African HTLV-II isolate, HTLV-II Gab, has been determined and analysed. Overall nucleotide sequence comparison showed that Gab is closer to the subtype b sequences (NRA, G12 and Gu) than to the prototypic subtype a sequence (Mo). This was confirmed by a detailed nucleotide and amino acid sequence comparison for each gene.

Several molecular features found in Gab clearly designated this isolate as an HTLV-II subtype b.

1. Nucleotide sequence analysis of the LTR revealed the presence of DraII (at nucleotide position 133) and AvaII (at nucleotide position 706) sites, both specific for subtype b. More detailed RFLP studies indicated that Gab presented a restriction profile close to, but slightly different from, the b5 profile described by Switzer *et al.* (1995b) and identical to the b3 profile described by Eiraku *et al.* (1995).

2. Hall *et al.* (1992) used a XhoI site, present in the nucleotide sequence of the envelope transmembrane gp21 of the Mo isolate and absent from the nucleotide sequence of NRA, G12 and Gu. The Gab isolate presented a C→T mutation at nucleotide 1030, thus disrupting the XhoI restriction site.

3. We observed the same nucleotide substitution at position 97 from the start of the pol gene, already observed with subtype b isolates. This resulted in the insertion of a stop codon at amino acid 32. A second ribosomal frameshift occurs downstream from codon 32 so that the presence of a stop codon at this position has no effect upon production of a functional polymerase protein.

(iv) Pardi *et al.* (1993a) observed an extended Tax protein in the subtype b isolates. The tax gene of our Gab isolate presented a change of two nucleotides towards the 3’ end of the protein, inserting an arginine residue for a stop codon normally present at amino acid 332 of the Mo Tax protein. This change resulted in a Tax protein 25 amino acids longer than the prototypic Mo Tax protein, giving it approximately the same size as HTLV-I-ATK. Eiraku *et al.* (1996) have determined the ability of shorter or longer Tax proteins to transactivate the HTLV-II LTR, using transient expression systems with chloramphenicol acetyltransferase (CAT) as a reporter gene. They found that the extended Tax protein had a much higher transactivation activity and that the carboxy terminus of Tax was required for effective transactivation. Differences between these longer and shorter Tax proteins have to be further studied to eventually correlate the type of Tax with virus phenotype.

The phylogenetic analysis based both on the env gp21 and on most of the LTR confirmed the presence of Gab within cluster b. According to studies on the LTR (Switzer *et al.*,...
African HTLV-II subtype b isolate

Fig. 4. Phylogenetic analysis of the LTR region (located between nucleotide positions 86–700) of different HTLV-II isolates using the maximum likelihood method (see Methods). Major branches (numbered 1–8) had $P$ values $<0.01$. All other branches of the tree had $P$ values of $<0.05$ to $<0.01$. The scale bar shows the ratio of nucleotide substitutions for a given horizontal branch length. Phylogroups defined by Switzer et al. (1995b) are indicated on the right of the tree.

1995b), Gab was shown to be present in phylogroup BIII close to the pygmy PYGCAM isolate. The restriction profiles presented by PYGCAM (b5 according to Switzer et al., 1995b) and Gab were shown to be very closely related, thus correlating our phylogenetic and RFLP results.

We observed a surprisingly close relationship between Gab and North American isolates: indeed, Gab was shown to be very close to NRA, to belong to phylogroup BIII with isolates of various North American origins, and to present a restriction profile b3 (Eiraku et al., 1995), previously identified in North American isolates. We confirmed the results obtained with PYGCAM (Gessain et al., 1995), thus pointing out the problem of the very low nucleotide divergence observed between African and American isolates compared with their very probable extremely long period of independent evolution (estimated between 50,000 and 100,000 years).

We compared the Env proteins of the available African subtype b isolates. Based on the Env amino acid sequence, Gab was closer to PYGCAM. Surprisingly, Gab was different from JPS, the other Gabonese isolate. Indeed, to date, two different subtype b isolates have been identified in Gabon. They originated from different regions (Gab from the north-west and JPS from the south-east), and from persons without any known relationship with the Americas and living in isolated areas. The presence of these two different isolates in Gabon and the close genetic proximity of Gab with the Cameroonian pygmy PYGCAM isolate strongly argue for an ancient presence of HTLV-II in Africa (Gessain & de The, 1996).

In view of the interspecies transmission of simian T-cell lymphotropic virus type I (STLV-I) (Koralnik et al., 1994; Liu et al., 1996) to humans in Central Africa and the recent evidence of new African primate T-cell lymphotropic viruses (Liu et al., 1994; Giri et al., 1994; Goubau et al., 1994; Van Brussel et al., 1996; Vandamme et al., 1996), our findings suggest the presence of a virus very closely related to HTLV-II subtype b in simian populations living or having lived in this Central Africa area. Future sero-epidemiological and molecular studies on these populations will possibly confirm the hypothesis of the African origin of the HTLV-II/STLV-II family.

We thank Sentob Saragosti for his precious help in the construction of the phylogenetic trees.

This study was supported by a Grant from l’Agence Nationale de Recherches sur le SIDA (ANRS).

References


African HTLV-II subtype b isolate


Received 3 July 1997; Accepted 1 October 1997