Journal of General Virology (1998), **79**, 269–277. Printed in Great Britain

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 Gessain[5] and Nicole Monplaisir[1]

[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'Epidémiologie des Virus Oncogenes, 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 epidemo-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 Méchain, 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.*, 1996; 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 (GhKt; Igarashi *et al.*, 1993) and Cameroon (PH230PCAM; Mauclère *et al.*, 1995).

2. Three HTLV-II subtype b isolates have been partially characterized: the first originated from plasma of a Zairean patient collected in 1969 (Dube *et al.*, 1994). The second originated from a Cameroonian pygmy of the Bakola tribe (PYGCAM; Gessain *et al.*, 1995) and was shown to be very close to the Amerindian 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 proteinase 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 $MgCl_2$, 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), AGCCACCCAGGGCGAGTCAT, and 202 (bp 468–487), CTAAGGGGCAGCCGAGCTCG; (2) 205 (bp 908–927),GCAGCCTAGGCCCTCCGATT, and 206 (bp 1521–1540), AGACCTTGCTGGGCGGGGTT; (3) 203 (bp 417–436), GGCCTCG-GCACCTCCTGAAC, and 204 (bp 1100–1119), TAGGTGTCGGAA-CTGGGGCG; (4) 207 (bp 1491–1508), ATGGCAGGGCCCCTAAGA, and 208 (bp 1959–1978), TCTTGGCATAGGGGGCAGGG; (5) 209 (bp 1879–1898), CCCCCACACAGCCCTGCTTT, and 210N (bp 2156–2175), GGTTGCTGTTGCTGCCGCAG; (6) 20.4BIS (bp 2040–2060), CTCCTGTTGGATCTCCCTTCC, and 210BIS (bp 2439–2459), GGGCGTCCCTTCCAATGATGG; (7) 239 (bp 2381–2400), TTCCG-AAGGTCCCCCGTTAT, and 22.2 (bp 3744–3763), GGTGGTGAA-TGAGGATGCCG; (8) 19.4 (bp 2973–2992), CCCCACCCTCTTC-GAACAAC, and 212 (bp 3464–3483), AGGTAGGGCGGGGTT-GAGGC; (9) 22.1 (bp 3705–3721), CCTTTGCGACTTCCTGA and 214 (bp 4053–4072), TAAGGGAAGGCCATGGCTTG; (10) OLIGO1 (bp 3995–4013), AAACACATTCCGCACAAAA and 31.2 (bp 4342–4361), GCGCCAAAAGAGACTAGAGC; (11) 217 (bp 4522–4541), TGCC-TCCACGTCTGGGTAGA and 219R (bp 5130–5149), CGAGTTCG-CTGGCGTCTGCT; (12) 31.1 (bp 4282–4301), GCTCCCCTTGTTC-CCCTGAC and 218NEW (bp 5019–5038), GGGATTGCAATGGA-CCTTTC; (13) 219 (bp 4989–5008), TAAACTCCCCGGCCTTACCA and 218BIS (bp 5228–5252), GTGTGCATCGGCTCTGCTGGG; (14) 25.1BIS (bp 5253–5272), TCACGATTGGTATCTCCTCC and 25.1BISREV (bp 5409–5428), ATGTGGGAATAAGTATAA; (15) 25.1 (bp 5232–5252), CCCAGCAGAGCCGATGCACAC and 220BIS (bp 5478–5497), TAGCGAGCAAGGGTCATTGT; (16) 26.5′ (bp 5469–5489), CGCCTTCCTACAATGACCCTT and 222 (bp 5993–6012), TGTAGGCGAGGTTGGTAGCA; (17) FL34S (bp 6284–6303), CAGTATGCAGCCCAAAATAG and 223REV (bp 6696–6715), GGTATAGAGGACTGTGGATG; (18) 223BIS (bp 5913–5932),

AACAAACCTCCTCCCGAACC and 224BIS (bp 6476–6492), GCTTCTCGTGCCCATTG; (19) 226REV (bp 6774–6793), TTCCTCTAACCCCCGCTCAC and 226 (bp 6968–6987), TTCTGCAGGAGCGTGAGGAG; (20) 235 (bp 6704–6723), GTCCTCTATACCAGATGAGT and 227BIS (bp 6820–6839), TTATGTGGATTTCCTGGAAGG; (21) 227 (bp 6907–6926), CTTCCCTCTCCCGGCGCTTT and 227REV (bp 7247–7266), CACGTAGACGGGGTATCCAT; (22) 228REV (bp 7144–7163), GGTCTCCTAACGGCAATCTC and 228 (bp 7616–7635), GTTTTGGGGACGGAGGCCAG; (23) 229BIS (bp 7591–7610), CTCCCCTCCCTCGCCTTCCC and 228BIS (bp7682–7701), TGTCATGGGTGGGGAAAGCT; (24) 229BIS (bp 7591–7610), CTCCCCTCCCTCGCCTTCCC and 229REV (bp 7918–7937), GTTGTTAAGATGGAGTGATA; (25) 230REV (bp 7855–7874), ACAATGTTCCAACCCGTGAG and 230BIS (bp 8253–8272), GATTGTTTGTGTGAGACGGT.

■ **Molecular cloning and sequencing.** PCR fragments were purified from low melting point agarose gel, kinased with T4 polynucleotide kinase and ligated to *Sma*I-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 46 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 Mo 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, Msa1Bp, 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 (Mauclère *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.RB, ITA47A and 50A, PENN7A, 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 (Mauclère *et al.*, 1995), PYGCAM1 (Gessain *et al.*, 1995), GhKt (Igarashi *et al.*, 1993), MEX17, KAY73 and 139 (Switzer *et al.*, 1996).

Sequence alignments were 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 Treepol 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 nucleotides 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 divergence).

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 *Dra*II 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 sequence available from the two other African HTLV-IIb isolates known to date with the two prototypic subtype b

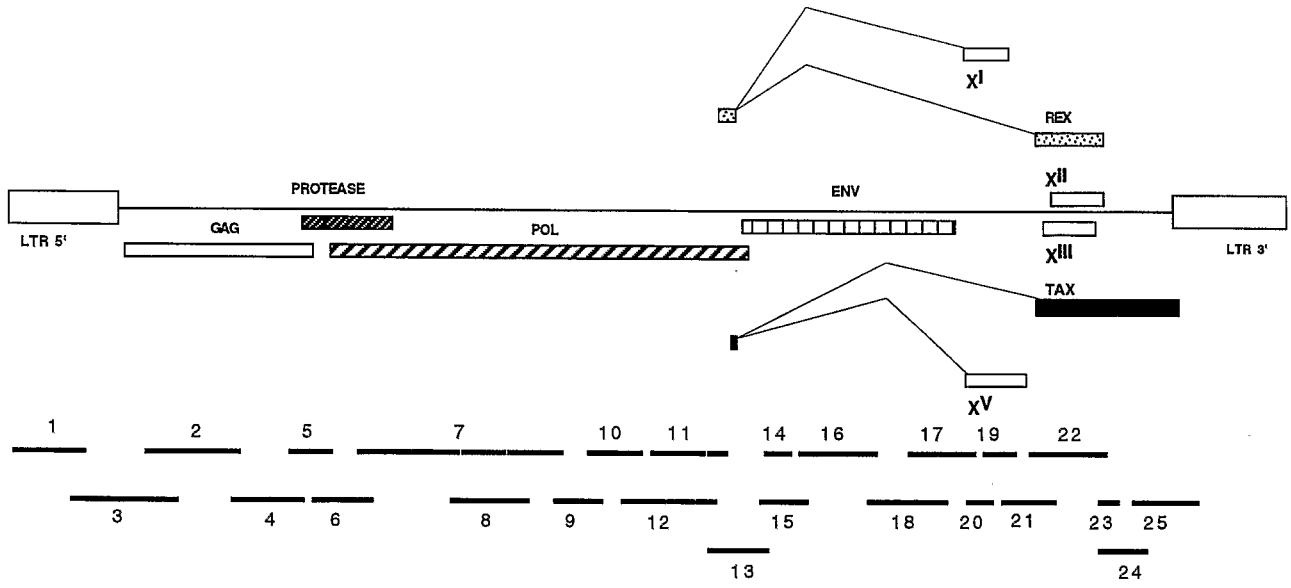

Fig. 1. Schematic representation of the strategy used to obtain the complete nucleotide sequence of HTLV-II-Gab. The positions of the structural *(gag, protease, pol and env)* and regulatory (*tax* and *rex*) genes, and of the five major open reading frames found in the pX region, are indicated by boxes.

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 (%)

HTLV-II-Gab compared with:	Overall	LTR	*gag*	*pol*	*env*	*pX*
HTLV-II-Mo	95·1	93·08	95·7	95·12	95·48	95·80
HTLV-II-NRA	99·3	98·83	99·31	99·40	98·97	99·42
HTLV-II-G12	98·9	98·30	98·92	99·25	98·97	99·35
HTLV-II-Gu	98·2	98·17	97·77	97·77	98·9	98·26

(b) Amino acid sequence similarity (%)

HTLV-II-Gab compared with:	*gag*	*pol*	*env*	*tax*	*rex*
HTLV-II-Mo	98·38	95·01	97·73	97·47	94·12
HTLV-II-NRA	100	99·49	98·55	99·16	99·41
HTLV-II-G12	99·77	99·18	98·77	99·16	99·41
HTLV-II-Gu	98·61	97·15	98·77	97·19	97·06

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 pygmy 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-IIb 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.* (1993*b*) 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.*, 1993*a*).

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: $p10^{XI}$,

Table 2. Amino acid sequence comparison of the *env*-encoded protein of HTLV-II-Gab with those of other HTLV-II isolates

	Amino acid divergence (%)					
	PYGCAM	JPS	G12	Gu	NRA	Mo
HTLV-II-Gab (Gabon)	1·44	2·47	1·23	1·23	1·45	2·27
PYGCAM (Cameroon)		2·26	1·02	0·62	1·23	2·67
JPS (Gabon)			1·65	1·65	2·26	3·08

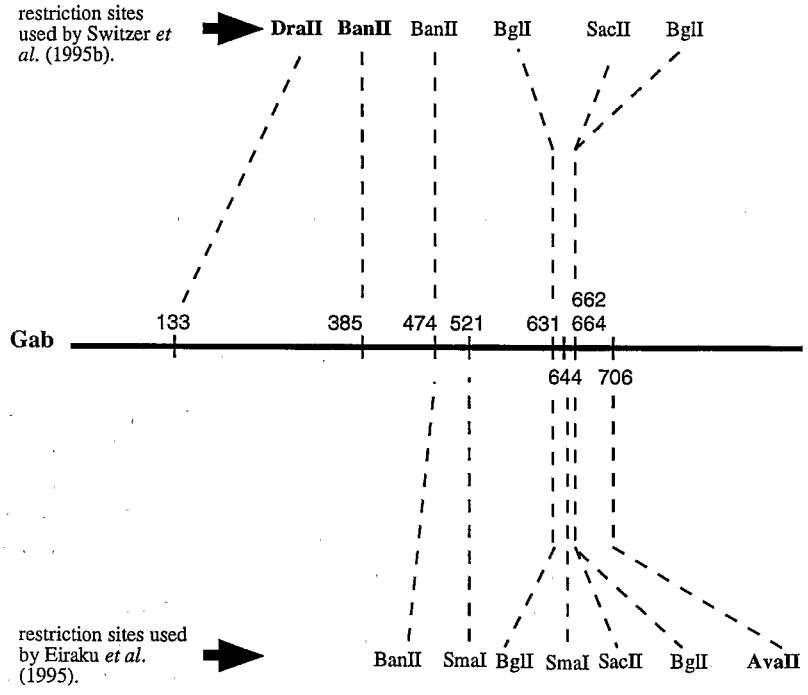

Fig. 2. Schematic representation of the RFLP analysis of the LTR. The upper and lower parts of the figure show the positions of restriction sites identified in HTLV-II-Gab, and used respectively by Switzer *et al.* (1995*b*) and Eiraku *et al.* (1995) to classify the different HTLV-II isolates. Nucleotide positions indicated for each restriction site are based on the Mo sequence. The *Ban*II site located at nucleotide position 385 was specific to HTLV-II-Gab. The *Dra*II (nt position 133) and *Ava*II (nt position 706) are specific for subtype b.

p11^{XV}, p22–20^{XIII} and p28^{XII}. Interestingly, a C → T mutation introduced a stop codon at amino acid 100 of p28^{XII} protein of NRA, G12 and Gab. The stop codon was absent in p28^{XII} of Mo and Gu, giving rise to a protein 117 amino acids longer.

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 *Xho*I 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, *Dra*II and *Ava*II 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.* (1995*b*) 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 *Dra*II and *Ava*II 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 *Ban*II 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

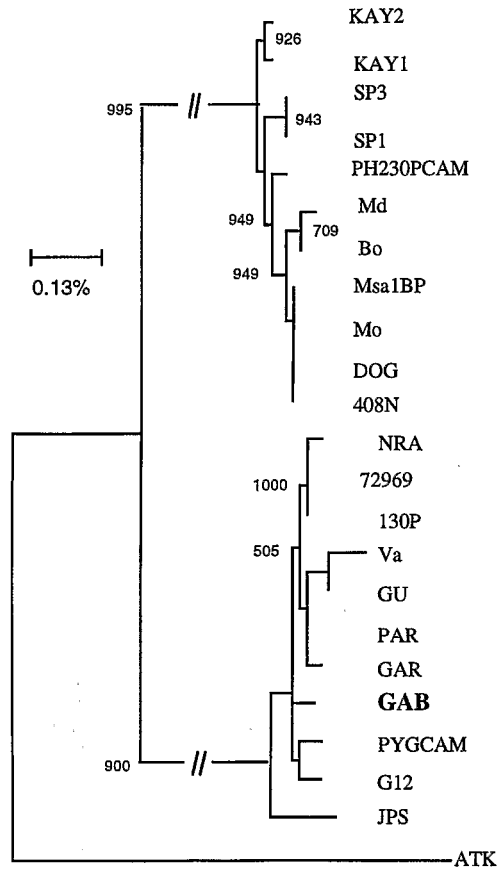

Fig. 3. Phylogenetic analysis of 589 nucleotides (located between nucleotide positions 6052–6640) of the *env* gene region encoding transmembrane glycoprotein gp21 of HTLV-II isolates, using the neighbor-joining method (see Methods). Bootstrap statistical analysis was applied using 1000 bootstraps replicates with ATK as the outgroup. Bootstrap values lower than 500 are not indicated.

PYGCAM and G12 isolates. However, the bootstrap values were too low to support the idea that a subcluster of these three isolates might exist (Fig. 3). The ML method gave a similar tree with an identical branching order.

Phylogenetic analysis of the LTR sequences (between nucleotides 86 and 700) was done by the NJ and ML methods. The topologies of the trees obtained with both methods were similar, demonstrating a clear separation (with high statistical significance) between the two subtypes a and b clades (Fig. 4). The ML analysis showed *P* values for major branches ranging from $< 0{\cdot}05$ to $< 0{\cdot}01$ (thus indicating the relevance of the phylogenetic results). Furthermore, our analysis identified the four phylogroups proposed by Switzer *et al.* (1995 *b*) within HTLV-IIb, and termed BI–BIV. Gab clustered in the BIII phylogroup, close to the PYGCAM isolate. Only gp21 analysis by the NJ method (Fig. 3) and LTR analysis by the ML method (Fig. 4) are shown; analysis of gp21 by ML and the LTR by NJ did not yield additional information.

The presence of Gab in the subtype b cluster after phylogenetic analysis based both on *env* gp21 and LTR 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 *Dra*II (at nucleotide position 133) and *Ava*II (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.* (1995 *b*) and identical to the b3 profile described by Eiraku *et al.* (1995).

2. Hall *et al.* (1992) used a *Xho*I 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 *Xho*I 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.* (1993 *a*) 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.*,

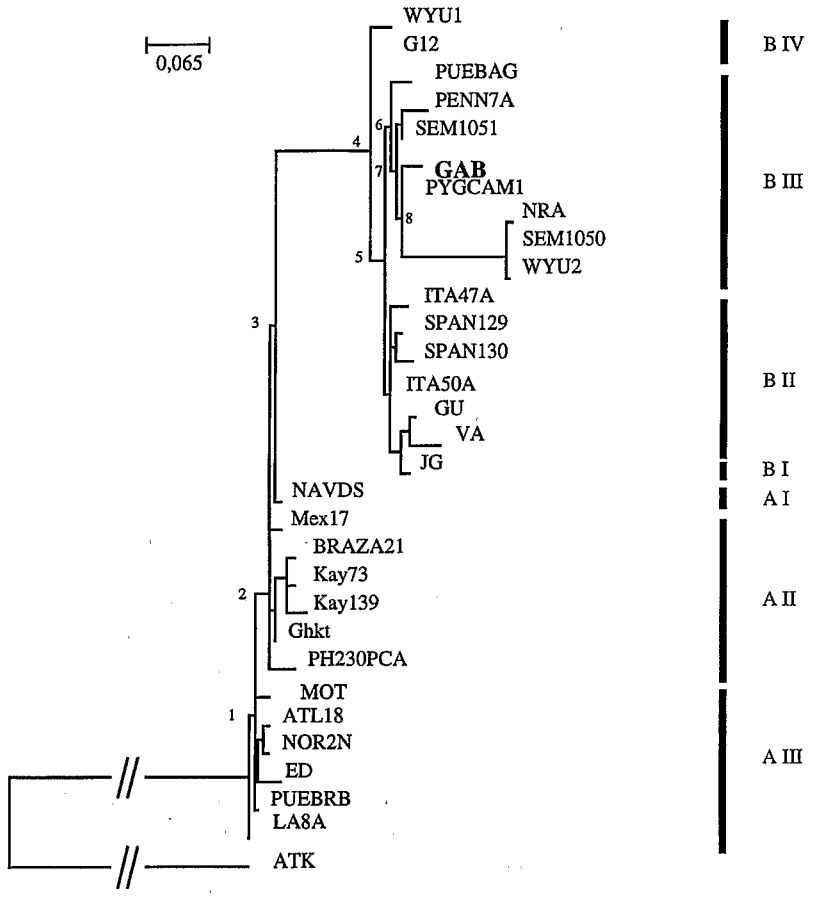

Fig. 4. Phylogenetic analysis of the LTR region (located between nucleotide positions 86–700) of different HTLV-II isolates using the maximum likelihood method (see Methods). Major branches (numbered 1–8) had P values $< 0{\cdot}01$. All other branches of the tree had P values of $< 0{\cdot}05$ to $< 0{\cdot}01$. The scale bar shows the ratio of nucleotide substitutions for a given horizontal branch length. Phylogroups defined by Switzer *et al.* (1995*b*) are indicated on the right of the tree.

1995*b*), 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.*, 1995*b*) 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 50000 and 100000 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

Ciminale, V., Pavlakis, G. N., Derse, D., Cunningham, C. P. & Felber, B. K. (1992). Complex splicing in the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and proteins produced by HTLV type I. *Journal of Virology* **66**, 1737–1745.

Ciminale, V., D'Agostino, D. M., Zotti, L., Franchini, G., Felber, B. K. & Chieco Bianchi, L. (1995). Expression and characterization of proteins produced by mRNAs spliced into the X region of the human T-cell leukemia/lymphotropic virus type II. *Virology* **209**, 445–456.

Delaporte, E., Monplaisir, N., Louwagie, J., Peeters, M., Martin-Prevel, Y., Louis, J.-P., Trebuco, A., Bedjabaga, L., Ossari, S., Honore, C., Larouze, B., d'Auriol, L., Van der Groen, G. & Piot, P. (1991). Prevalence of HTLV-I and HTLV-II infection in Gabon, Africa: comparison of the serological and PCR results. *International Journal of Cancer* **49**, 373–376.

Dube, D. K., Dube, S., Erensoy, S., Jones, B., Bryz Gornia, V., Spicer, T., Love, J., Saksena, N., Lechat, M. F., Shrager, D. I. and others (1994). Serological and nucleic acid analyses for HIV and HTLV infection on archival human plasma samples from Zaire. *Virology* **202**, 379–389.

Eiraku, N., Monken, C., Kubo, T., Zhu, S. W., Rios, M., Bianco, C., Hjelle, B., Nagashima, K. & Hall, W. W. (1995). Nucleotide sequence and restriction fragment length polymorphism analysis of the long terminal repeat of human T cell leukemia virus type II. *AIDS Research and Human Retroviruses* **11**, 625–636.

Eiraku, N., Novoa, P., Costa Ferreira, M., Monken, C., Ishak, R., Costa Ferreira, O., Zhu, S. W., Lorenco, R., Ishak, M., Azvedo, V., Guerreiro, J., Pombo de Oliveira, M., Loureiso, P., Hammerschlak, N., Ijishi, S. & Hall, W. W. (1996). Identification and characterization of a new and distinct molecular subtype of human T-cell lymphotropic virus type 2. *Journal of Virology* **70**, 1481–1492.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Fouchard, N., Flageul, B., Bagot, M. and others (1995). Lack of evidence of HTLV-I/II infection in T CD8 malignant or reactive lymphoproliferative disorders in France: a resological and/or molecular study of 169 cases. *Leukemia* **9**, 2087–2092.

Gessain, A. & de The, G. (1996). What is the situation of human T cell lymphotropic virus type II (HTLV-II) in Africa? Origin and dissemination of genomic subtypes. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **13**(S1), S228–235.

Gessain, A., Vernant, J. C., Maurs, L., Barin, F., Gout, O., Glender, A. & de The, G. (1985). Antibodies to human T-lymphotropic virus type 1 in patients with tropical spastic paraparesis. *Lancet* ii, 407–409.

Gessain, A., Mauclère, P., Froment, A., Biglione, M., Le Hesran, J. Y., Tekaia, F., Millan, J. & de The, G. (1995). Isolation and molecular characterization of a human T-cell lymphotropic virus type II (HTLV-II), subtype B, from a healthy Pygmy living in a remote area of Cameroon: an ancient origin for HTLV-II in Africa. *Proceedings of the National Academy of Sciences, USA* **92**, 4041–4045.

Giri, A., Markham, P., Digilio, L., Hurteau, G., Gallo, R. C. & Franchini, G. (1994). Isolation of a novel simian T-cell lymphotropic virus from *Pan paniscus* that is distantly related to the human T-cell leukemia/lymphotropic virus types I and II. *Journal of Virology* **68**, 8392–8395.

Goubau, P., Van Brussel, M., Vandamme, A.-M., Liu, H. F. & Desmyter, J. (1994). A primate T-lymphotropic virus, PTLV-L, different from human T-lymphotropic viruses types I and II, in a wild-caught baboon (*Papio hamadrias*). *Proceedings of the National Academy of Sciences, USA* **91**, 2848–2882.

Hall, W. W., Takahashi, H., Liu, C., Kaplan, M. H., Scheewind, O., Ijichi, S., Nagashima, K. & Gallo, R. C. (1992). Multiple isolates and characteristics of human T-cell leukemia virus type II. *Journal of Virology* **66**, 2456–2463.

Hall, W. W., Ishak, R., Zhu, S. W., Novoa, P., Eiraku, N., Takahashi, H., da Costa Ferreira, M., Azevedo, V., Ishak, M. O. G., da Costa Ferreira, O., Monken, C. & Kurata, T. (1996). Human T lymphotropic virus type II (HTLV-II): epidemiology, molecular properties and clinical features of infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **13**(S1), S204–214.

Heneine, W. (1996). The phylogeny and molecular epidemiology of human T cell lymphotropic virus type II. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **13**(S1), S236–241.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences* **8**, 189–191.

Hjelle, B., Zhu, S. W., Takahashi, H., Ijichi, S. & Hall, W. W. (1993). Endemic human T cell leukemia virus type II infection in southwestern US Indians involves two prototype variants of virus. *Journal of Infectious Diseases* **168**, 737–740.

Igarashi, T., Yamashita, M., Miura, T., Osei Kwasi, M., Aysi, N. K., Shiraki, H., Kurimura, T. & Hayami, M. (1993). Isolation and genomic analysis of human T lymphotropic virus type II from Ghana. *AIDS Research and Human Retroviruses* **9**, 1039–1042.

Ishak, R., Harrington, W. J., Azevedo, V. N., Eiraku, N., Ishak, M. O. G., Guerreiro, J. F., Santos, S. B., Kubo, T., Monken, C., Alexander, S. & Hall, W. W. (1995). Identification of human T cell lymphotropic virus type IIa infection in the Kayapo, an indigenous population of Brazil. *AIDS Research and Human Retroviruses* **11**, 989–993.

Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, M., Golde , D. & Gallo, R. C. (1982). A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science* **218**, 571–573.

Koralnik, I. J., Boeri, E., Saxinger, W. C. and others (1994). Phylogenetic associations of human and simian T-cell leukemia/lymphotropic virus type I strains: evidence for interspecies transmission. *Journal of Virology* **68**, 2693–2707.

Lee, H., Idler, K. B., Swanson, P., Aparicio, J. J., Chin, K. K., Lax, J. P., Nguyen, M., Mann, T., Leckie, G., Zanetti, A., Marinucci, G., Chen, I. S. Y. & Rosenblatt, J. D. (1993). Complete nucleotide sequence of HTLV-II isolate NRA: comparison of envelope sequence variation of HTLV-II isolates from U.S. blood donors and U.S. and Italian i.v. drug users. *Virology* **196**, 57–69.

Lipka, J. J., Miyoshi, I., Hadlock, K. G., Reyes, G. R., Chow, T. P., Blattner, W. A., Hanson, C. V., Gallo, D., Chan, L. & Foung, S. K. H. (1992). Segregation of human T cell lymphotropic virus type I and II infections by antibody reactivity to unique viral epitopes. *Journal of Infectious Diseases* **165**, 268–272.

Liu, H. F., Vandamme, A. M., Van Brussel, M., Desmyter, J. & Goubau, P. (1994). New retroviruses in human and simian T-lymphotropic viruses. *Lancet* **344**, 265–266.

Liu, H. F., Goubau, P., Van Brussel, M., Van Laethem, K., Chen, Y.-C., Desmyter, J. & Vandamma, A.-M. (1996). The three human T-lymphotropic virus type I subtypes arose from three geographically distinct simian reservoirs. *Journal of General Virology* **77**, 359–368.

Mauclère, P., Mahieux, R., Garcia-Calleja, J. M., Salla, R., Tekaia, F., Millan, J., de The, G. & Gessain, A. (1995). A new HTLV-II subtype a isolate in an HIV-1 infected prostitute from Cameroon, Central Africa. *AIDS Research and Human Retroviruses* **11**, 989–993.

Pardi, D., Kaplan, J. E., Coligan, J. E., Folks, T. M. & Lal, R. B. (1993*a*). Identification and characterization of an extended Tax protein in human T-cell lymphotropic virus type II subtype b isolates. *Journal of Virology* **67**, 7663–7667.

Pardi, D., Switzer, W. M., Hadlock, K. G., Kaplan, J. E., Lal, R. B. & Folks, T. M. (1993*b*). Complete nucleotide sequence of an Amerindian human T-cell lymphotropic virus type II (HTLV-II) isolate: identification

of a variant HTLV-II subtype b from a Guaymi Indian. *Journal of Virology* **67**, 4659–4664.

Poiesz, B., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna , J. D. & Gallo, R. C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences, USA* **77**, 7415–7419.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Salemi, M., Cattaneo, E., Casoli, C. & Bertazzoni, U. (1995). Identification of IIa and IIb molecular subtypes of human T-cell lymphotropic virus type II among Italian injecting drug users. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **8**, 516–520.

Salemi, M., Vandamme, A.-M., Guano, F., Gradozzi, C., Cattaneo, E., Casoli, C. & Bertazonni, U. (1996). Complete nucleotide sequence of the Italian human T-cell lymphotropic virus type II isolate Gu and phylogenetic identification of a possible origin of South European epidemics. *Journal of General Virology* **77**, 1193–1201.

Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proceedings of the National Academy of Sciences, USA* **80**, 3618–3622.

Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M. & Sugimura, T. (1985). Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. *Proceedings of the National Academy of Sciences, USA* **82**, 3101–3105.

Switzer, W. M., Owen, S. M., Pieniazek, D., Nerurkar, V., Duenas-Barajas, E., Heneine, W. & Lal, R. B. (1995*a*). Molecular analysis of human T-cell lymphotropic virus type II from Wayu indians of Colombia demonstrates two subtypes of HTLV-IIb. *Virus Genes* **10**, 153–162.

Switzer, W. M., Pieniazek, D., Swanson, P., Samdal, H. H., Soriano, V., Khabbaz, R. F., Kaplan, J. E., Lal, R. B. & Heneine, W. (1995*b*). Phylogenetic relationship and geographic distribution of multiple human T-cell lymphotropic virus type II subtypes. *Journal of Virology* **69**, 621–32.

Switzer, W. M., Black, F. L., Pieniazek, D., Biggar, R. J., Lal, R. B. & Heneine, W. (1996). Endemicity and phylogeny of the human T cell lymphotropic virus type II subtype A from the Kayapo Indians of Brazil: evidence for limited regional dissemination. *AIDS Research and Human Retroviruses* **12**, 635–640.

Takahashi, H., Zhu, S. W., Ijichi, S., Vahlne, A., Suzuki, H. & Hall, W. W. (1993). Nucleotide sequence analysis of human T cell leukemia virus, type II (HTLV-II) isolates. *AIDS Research and Human Retroviruses* **9**, 721–732.

Tuppin, P., Gessain, A., Kazanji, M., Mahieux, R., Cosnefroy, J.-Y., Tekaia, F., Georges-Courbot, M.-C., Georges, A. & de The, G. (1996). Evidence in Gabon for an intrafamilial clustering with mother-to-child and sexual transmission of a new molecular variant of human T-lymphotropic virus type-II subtype B. *Journal of Medical Virology* **48**, 22–32.

Van Brussel, M., Goubau, P., Rousseau, R., Desmyter, J. & Vandamme, A.-M. (1996). The genomic structure of a new primate T-lymphotropic virus, STLV-PH969, differs from that of simian T-lymphotropic virus type I and human T-lymphotropic virus type I and II. *Journal of General Virology* **77**, 347–358.

Vandamme, A.-M., Liu, H.-F., Van Brussel, M., De Meurichi, W., Desmyter, J. & Goubau, P. (1996). The presence of a divergent T-lymphotropic virus in a wild-caught pygmy chimpanzee (*Pan paniscus*) supports an African origin for the human T-lymphotropic/simian T-lymphotropic group of viruses. *Journal of General Virology* **77**, 1089–1099.

Received 3 July 1997; Accepted 1 October 1997