Phylogenetic classification of human T cell leukaemia/lymphoma virus type I genotypes in five major molecular and geographical subtypes

Abel Ureta Vidal,1 Antoine Gessain,1 Mitsuyaki Yoshida,2 Fredj Tekaia,3 Benoît Garin,1 Bernard Guillema,4 Thomas Schulz,5 Reza Farid6 and Guy De Thé†∗

1 Epidemiology of Oncogenic Viruses, Pasteur Institute, 28 rue du Dr Roux, 75724 Paris Cedex 15, France, 2 The Institute of Medical Science, Tokyo University, 4-6-1 Shirokanedai Minato-ku Tokyo 108, Japan, 3 Service d’Informatique Scientifique, Pasteur Institute, Paris, France, 4 INSERM 328, Fondation Bergonié, 229 Cours de l’Argonne, 33076 Bordeaux, France, 5 Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London W3 6JB, U.K. and 6 Mashad University of Medical Sciences, Mashad, Iran.

Proviral DNA was obtained from ex vivo peripheral blood mononuclear cells of 75 human T cell leukaemia/lymphoma virus type I (HTLV-I)-infected individuals who were either asymptomatic or had adult T cell leukaemia or tropical spastic paraparesis/HTLV-I-associated myelopathy. Amplified long terminal repeats (LTRs) were analysed for restriction fragment length polymorphisms (RFLPs). The results, together with previously published LTR data (a total of 180 specimens analysed), showed the presence of 12 different RFLP profiles with four major molecular subtypes. Furthermore, a fragment of 413 bp (nucleotides 22 to 434) of the U3/R region was sequenced for 12 new HTLV-I specimens originating from Central and West Africa (8 cases), Iran (1 case), Caribbean (2 cases) and Reunion Island (1 case). Phylogenetic analysis using three different techniques (maximum parsimony, neighbour-joining and UPGMA) comparing these 12 strains (including four new African HTLV-I variants) with the 30 published partial HTLV-I LTR sequences (nt 120 to 434) showed the existence of clusters of molecular variants in discrete geographical areas. The topology of the phylogenetic trees is thought to reflect HTLV-I evolution and the migrations of virally infected populations in the recent or distant past. Furthermore, there was a nearly perfect concordance between the clustering based on the LTR sequence homologies and the LTR RFLP subtypes suggesting that this rapid and simple technique is well suited to the investigation of HTLV-I molecular epidemiology. These results allow a new phylogenetic classification of HTLV-I genotypes into five major molecular subtypes: Cosmopolitan (C) subtype widespread all over the world, Japanese (J) subtype, West African (WA) subtype, Central African (CA) subtype and Melanesian (M) subtype.

Human T cell lymphotropic virus type I (HTLV-I), the first human oncoretrovirus isolated (Poiesz et al., 1980, 1993; Miyoshi et al., 1981) is the aetiological agent of the CD4+ lymphoproliferative disease adult T cell leukaemia (ATL) (Takatsuki et al., 1977; Hinuma et al., 1982) characterized by a clonal integration of HTLV-I provirus in the tumour cells (Yamaguchi et al., 1984; Yoshida et al., 1982, 1984) and of a chronic neuro-myelopathy, tropical spastic paraparesis/HTLV-I-associated myelopathy (TSP/HAM) (Gessain et al., 1985; Osame et al., 1986; Gout et al., 1990). HTLV-I is endemic in tropical Africa, south-eastern Japan, the Caribbean area, Central and South America and some regions of Melanesia (Hinuma et al., 1982; De Thé et al., 1985; Blattner, 1989; Yanagihara et al., 1991). In these regions, between 0.5 and 20% of the general population, depending on age and sex, have HTLV-I antibodies and are considered to be healthy HTLV-I carriers. The question whether the same HTLV-I strain could induce several diseases through different pathways, as in the EBV system (De Thé, 1993), or whether there are specific gene mutations that direct tissue tropism and pathogenesis, as in the case of murine leukaemia retroviruses (Desgroseillers et al., 1985; Rassart et al., 1986; Li et al., 1987; Szurek et al., 1988), led to the comparative molecular analysis of different HTLV-I viral strains (Daenke et al., 1990; De et al., 1991; Komurian et al., 1991, 1992a, b; Kinoshita et al., 1991; Ratner et al., 1991; Schulz et al., 1991; Paine et al., 1991; Ehrlich et al., 1992; Saksena et al., 1992; Sherman et al., 1992). These studies (performed mostly on the LTR and part of the pol and env gene of samples originating from Japan, the Caribbean, South America and the U.S.A.) led to the conclusion that the low in vivo genomic variability of HTLV-I depends much more upon geography than pathology (Komurian et al., 1991;
Kinoshita et al., 1991; Ratner et al., 1991; Gessain et al., 1992a). Furthermore, the puzzling epidemiological distribution of HTLV-I with the existence of geographical molecular clusters, and the recent discovery of distant molecular variants of HTLV-I in Central Africa (Ratner et al., 1985; Gessain et al., 1992a, b; Boeri et al., 1993) and Melanesia (Gessain et al., 1991, 1993; Sherman et al., 1992; Saikena et al., 1992; Bastian et al., 1993) raised new questions concerning the origin and evolution of this human oncoretrovirus.

In order to gain new insights on the global dissemination of HTLV-I and the phylogenetic relationship between the different HTLV-I strains, we have performed a molecular analysis of a large variety of HTLV-I proviral DNAs from patients with ATL, TSP/HAM or healthy HTLV-I carriers and originating from most of the known HTLV-I endemic areas. The LTR region was selected because this region, especially the U3 fragment, is the most variable of the HTLV-I genome (Daenke et al., 1990; Komurian et al., 1991), and also because we developed an RFLP technique (Komurian-Pradel et al., 1992) allowing the rapid study of nucleotide mutations in this portion of the genome. Finally LTR sequences of other HTLV-I isolates have been previously reported (Ratner et al., 1991; Ureta Vidal et al., 1994) allowing phylogenetic comparison.

In a first step, using RFLP analysis we studied 75 new proviral DNA samples, mostly extracted from ex vivo peripheral blood mononuclear cells (PBMC) specimens for PCR. Primers LTR1 and LTR2 (Komurian-Pradel et al., 1992) were used to amplify the LTR region of HTLV-I located between nucleotides 9 and 746. Briefly, 1 μg of ex vivo PBMC DNA was used for each PCR and 35 cycles were performed using Taq DNA polymerase and a DNA Thermocycler (Perkin Elmer Cetus) with denaturation at 94°C for 1 min, annealing at 58°C for 1 min and extension at 72°C for 2 min (increasing by 2 s per cycle). Ten μl of the amplified product (738 bp long) was analysed on 1.4% agarose gel and was then transferred to a nylon filter and hybridized with the 32P-labelled oligonucleotide LTR3 probe (Komurian-Pradel et al., 1992) to confirm the specificity of the PCR product. One twentieth of purified amplified DNA was digested with 2 units of Apal, NdiI, MaeIII, Dral or SacI (Boehringer Mannheim), respectively. The digested DNA was separated on a 4% agarose gel, transferred to a nylon filter and hybridized with total LTR probe labelled with the ECL gene detection system (Amersham) based on enhanced chemiluminescence. Furthermore, the results from our preliminary RFLP study based on 95 DNA specimens (Komurian-Pradel et al., 1992; Ureta Vidal et al., 1994) and 10 previously published complete LTR sequences were also included in the final analysis of results. Thus a total of 180 different samples of HTLV-I obtained from healthy carriers or patients presenting with ATL or TSP/HAM were analysed. Table 1 presents clinical and geographical origin of these 180 specimens which cover all known endemic regions. Twelve different RFLP profiles were observed among these 180 HTLV-I specimens (see Table 2), thus showing a more complex pattern than described in our initial study (Komurian-Pradel et al., 1992). However, a simpler classification, obtained by grouping DNAs with common restriction profiles, led to four major molecular subtypes (Table 2). The Cosmopolitan subtype (formerly subtype II of Komurian-Pradel et al., 1992) was found to be widespread over the world (77/180 specimens) whereas the African subtype (formerly subtype I) (25/180) and the Japanese subtype (formerly subtype III) (77/180) were present in restricted areas. Finally the Melanesian subtype was observed solely in the South Pacific sample. There was no association between any of the RFLP subtypes and the clinical status of the patients (results not shown). In contrast, there was a striking clustering of the RFLP HTLV-I subtypes according to the geographical origin of the specimens (Table 3): most of the African specimens (22/25: 88%) included the new isolates from Mali, Gabon and Zaire, exhibited a common subtype, whereas most of the Caribbean and American samples (46/49: 94%) were of the Cosmopolitan subtype. Of interest, two subtypes were detected among the Japanese specimens including the eight Hawaiian samples of Japanese origin. Although 77 out of 99 (78%) shared the same RFLP profile named Japanese, 22/99 (22%) exhibited the Cosmopolitan profile. The five new samples from the Middle East (Iran), and that from Reunion Island (Indian Ocean), were of Cosmopolitan subtype.

In a second step, nucleic acid analysis by three different

<table>
<thead>
<tr>
<th>Clinical diagnosis</th>
<th>West &amp; Central Africa*</th>
<th>Caribbean/ Americas†</th>
<th>Middle East/ Reunion‡</th>
<th>Japan/ Pacific§</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptomatic carrier</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>48</td>
<td>65</td>
</tr>
<tr>
<td>ATL</td>
<td>7</td>
<td>19</td>
<td>4</td>
<td>23</td>
<td>53</td>
</tr>
<tr>
<td>TSP/HAM</td>
<td>9</td>
<td>21</td>
<td>1</td>
<td>28</td>
<td>59</td>
</tr>
<tr>
<td>Other†‡</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>49</td>
<td>6</td>
<td>100</td>
<td>180</td>
</tr>
</tbody>
</table>

* Includes specimens from Zaire (11), Ivory Coast (9), Gabon (1), Congo (1), Mali (1), Central African Republic (1) and Senegal (1).
† Includes specimens from French West Indies (23), Jamaica (1), Caribbean (2), Haiti (1), French Guiana (11), Brazil (6) and U.S.A. (5).
‡ Includes specimens from Iran (5) and Reunion (1).
§ Includes specimens from Japan (90), China (1), Hawaii (8) and Melanesia (1).
¶ Includes non-spastic paraparesis (1) B cell lymphoma (1) and acute lymphocytic leukaemia (1).
phylogenetic techniques was performed on 42 partial
LTR sequences including 12 new sequences generated in
the present work. These 12 specimens originated mostly
from newly described HTLV-I endemic areas including
ATL originating from Mali (DIA), Gabon (IGU), Iran (SAS)
and West Indies (pt18), as well
as TSP/HAM originating from Central African Republic
(KOP), Ivory Coast (KOU) and Zaire (MAS, Z15) or
healthy HTLV-I carriers from the French West Indies,
Martinique (BUL), Zaire (Z59) and Reunion Island
(MEL). The molecular data on the HTLV-I from Gabon,
Mali, Iran and Reunion Island represents the first
sequences obtained from these areas. For PCR,
primers P3LTR (5' TTTGAGCGGCCGCTGACAATGACCATGAGCCCCA 3') and LTRU5E (5' ACTTAGAATTCGCAGTTCAGGAGGACCACAGGCG 3')
(Gessain et al., 1993) were used to amplify all the U3
region and part of the R region of the LTR between
nucleotides 1 and 459. Amplification was performed
under the same conditions as with LTR1 and LTR2
primers. The PCR product was cloned and double-
stranded sequencing was done. Two clones of the 413 bp
fragment (nucleotides 22 to 434) from ex vivo specimens
of five of the 12 new HTLV-I isolates and one clone of
the 7 others were sequenced. The sequences of the LTR
regions known to have a functional importance were
highly conserved in these 12 new isolates and no
mutations were seen in the 21 bp imperfect repeat
enhancer elements except in the Zairian and Gabonese
strains exhibiting substitutions at positions 102 (A to G),
151 (A to G), 255 (G to A) for MAS, at positions 102 and
151 for Z15 and Z59, and only at 151 for IGU. However,
none of these mutations occurred in the consensus
sequence, TGACC, of the central core motif of these
enhancer elements known to respond to both cyclic
AMP and HTLV-I Tax-induced protein. Similarly, no
nucleotide change as compared to the HTLV-I ATK
prototype (Seiki et al., 1983) was observed in the poly(A)
signal (AATAAA), the TATA box and the CAP site.
Most of the mutations observed were single base

---

Table 2. Results of RFLP analysis of 180 HTLV-I proviral DNA samples showing 12 restriction profiles
classified into four major subtypes

<table>
<thead>
<tr>
<th>ApaI</th>
<th>NdeI</th>
<th>MaeIII</th>
<th>Dral</th>
<th>SacI</th>
<th>Previous subtype</th>
<th>Number of specimens</th>
<th>Proposed subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>I</td>
<td>17</td>
<td>African a</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>6 African b</td>
<td>6</td>
<td>African b</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>1 African c</td>
<td>1</td>
<td>African c</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>1 African d</td>
<td>1</td>
<td>African d</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>II</td>
<td>52</td>
<td>Cosmopolitan a</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13 Cosmopolitan b</td>
<td>13</td>
<td>Cosmopolitan b</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5 Cosmopolitan c</td>
<td>5</td>
<td>Cosmopolitan c</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2 Cosmopolitan d</td>
<td>2</td>
<td>Cosmopolitan d</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4 Cosmopolitan e</td>
<td>4</td>
<td>Cosmopolitan e</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>1 Cosmopolitan f</td>
<td>1</td>
<td>Cosmopolitan f</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>III</td>
<td>77</td>
<td>Japanese</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>1 Melanesian</td>
<td>1</td>
<td>Melanesian</td>
</tr>
</tbody>
</table>

* Symbols: '+' indicates the presence of a restriction enzyme site and '-' its absence, grey shading indicates features consistent
in a given subtype.
† According to the initial classification of Komurian et al. (1992).

Table 3. Distribution of HTLV-I proviral DNA specimens according to geographical origin and proposed subtype

<table>
<thead>
<tr>
<th>Total</th>
<th>West/Central Africa</th>
<th>Caribbean/ America</th>
<th>Middle East/ Reunion</th>
<th>Japan/ Hawaii</th>
<th>Melanesia</th>
<th>RFLP subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>African</td>
</tr>
<tr>
<td>77</td>
<td>3</td>
<td>46</td>
<td>6</td>
<td>22</td>
<td>0</td>
<td>Cosmopolitan</td>
</tr>
<tr>
<td>77</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>77</td>
<td>0</td>
<td>Japanese</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Melanesian</td>
</tr>
<tr>
<td>180</td>
<td>25</td>
<td>49</td>
<td>6</td>
<td>99</td>
<td>1</td>
<td>Total</td>
</tr>
</tbody>
</table>

---
Table 4. Nucleotide sequence alignment and comparison of a 413 bp region of the LTR of 42 HTLV-I sequences*

<table>
<thead>
<tr>
<th>Melanesian Central</th>
<th>West African</th>
<th>Cosmopolitan subtype</th>
<th>Japanese subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>subtype</td>
<td>African subtype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subtype</td>
<td>subtype</td>
<td></td>
<td>subtype</td>
</tr>
</tbody>
</table>

* The table contains a nucleotide sequence alignment comparing different subtypes of HTLV-I. The sequences are presented in a matrix format, with each row and column representing different subtypes and positions within the LTR. The nucleotide sequences are aligned and compared to highlight conserved and variant regions among the different subtypes.
### Table 4. (continued).

<table>
<thead>
<tr>
<th>Melanesian subtype</th>
<th>Central African subtype</th>
<th>West African subtype</th>
<th>Cosmopolitan subtype</th>
<th>Japanese subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central African</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West African</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosmopolitan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ATM sequence corresponds to the resequenced data in Ureta Vidal et al. (1994). On left and right the numbering according to the ATK prototype strain is shown (Seiki et al., 1983). Symbols: X, deletion; nd, not done. Substitutions or deletions specific for a particular subtype are boxed.

substitutions, although rare cases of insertion and deletion were also present. The four LTR sequences from Zairian and Gabonese strains were the most distant (6.2 to 5.3%) from the ATK prototype, whereas the other eight strains exhibited less than 3% nucleotide divergence from the HTLV-I ATK prototype in this LTR region.

We compared the DNA sequences of these 12 new LTRs with the 30 published total or partial LTR sequences from which RFLP subtypes were determined (Table 4). These latter 30 include 18 sequences from Japan ([Jap1, Jap2, Jap3, Jap4] (Komurian et al., 1991), H5 (Tsujimoto et al., 1988), HCT (Shirabe et al., 1990), ATK (Seiki et al., 1983), ATM (Seiki et al., 1982), DJ1, DJ9, DJ16, DJ29, DJ30, KM1, KM8, KG18, RK2, RK7 (Ureta Vidal et al., 1994)], three sequences from Africa [EL (Ratner et al., 1991), Sie, Akr (Komurian et al., 1991)]; six sequences from the Caribbean area [HS35 (Malik et al., 1988), CH (Ratner et al., 1991), TSP1 (Evangelista et al., 1990), Gro, Xav, Bou (Komurian et al., 1991)]; one sequence from the U.S.A., CR1 (Josephs et al., 1984); one sequence from Papua New Guinea, PNG-1 (Saksena et al., 1992) and one sequence from a Solomon Islander, MEL5 (Gessain et al., 1993). Alignments made on 413 sequenced base pairs of each LTR of the 42 proviral DNAs clearly demonstrated the existence of five molecular subtypes of HTLV-I, as defined by specific nucleotide substitutions or deletions (Table 4).

The first group (ATK to DJ9) formed the Japanese subtype (J), the second group (CH to pt18) formed the Cosmopolitan subtype (C), the third group (AKE to DIA) formed the West African subtype (WA) and the fourth group (EL to IGU) formed the Central African subtype (CA). Finally, the last group (PNG-1 and MEL5) formed the Melanesian subtype (M). These mutations can easily be used to trace the origin of a given proviral DNA. As an example, all the Cosmopolitan type strains have a G in position 306 and a deletion in 209, except for the CH isolate. The Central African group has an A substitution in positions 174 and 233 and a T substitution in positions 248 and 283 as compared to ATK. However, there are a few exceptions: the specimen

---

*Fig. 1. (See pages 3660 and 3661). Three phylogenetic trees were obtained after sequence alignment and bootstrapping, one by the maximum parsimony method (Fig. 1a), the second by the neighbour-joining method (data not shown) and the third one (Fig. 1b) by the UPGMA method applied to the mean distance matrix. The compared sequences correspond to a 315 base pair fragment of the LTR region encompassing most of the U3 and the R region. The numbers indicated at some nodes represent their frequency of occurrence out of 100 trees and therefore a measure of the robustness of the proposed trees. The elements of a given node are clustered together. In the tree of Fig. 1(a) the branch lengths are not proportional to nucleotide substitutions. In Fig. 1(b) the length of each branch is directly proportional to the expected number of substitutions that have occurred on that branch. The three phylogenetic trees correspond to 42 different HTLV-I isolates including the 12 new isolates generated in this study and the 30 other available published sequences. The HTLV-II MO isolate was used as outgroup to root the tree. The topology of the three trees was very similar, demonstrating the existence of five molecular clusters of HTLV-I genotypes.*
Fig. 1(a). For legend see page 3659.
Percentage of expected number of substitutions

Fig. 1(b). For legend see page 3659.
Japanese specimens; the West African cluster included the isolates from Ivory Coast, Mali and the only Jamaican isolate; the Central African cluster included all the Zairian and Gabonese isolates, and the Melanesian group included the two distant strains from Solomon Island and Papua New Guinea. The Melanesian, Central African and West African groups were quite distinct clusters on the maximum parsimony analysis (Fig. 1a), separated from the others in 100%, 84% and 78%, respectively, of the bootstrap results. The Japanese and the Cosmopolitan groups were not as stable (42%). Lower percentages were found with the NJ analysis for the African clusters (data not shown).

By phylogenetic tree topology, the concordance between geographical molecular clusters and LTR RFLP subtypes appeared nearly perfect. For example on both maximum parsimony analysis (Fig. 1a) and NJ tree (data not shown), the 20 samples classified in RFLP as Cosmopolitan subtype clustered together on the top of this tree based on sequence analysis. All the samples belonging to RFLP Japanese subtype clustered together. Similarly the 11 samples classified as RFLP African subtype clustered together, representing on the bottom of the tree a wide aggregate which could be subdivided into two groups, one of the West African specimens, and one of the Central African specimens. The sample from Mali (DIA; of RFLP African subtype), exhibited a unique phylogenetic position within the hierarchical order of the HTLV-Is, being placed between the Japanese and West African subtypes. In Fig. 1(b) the length of the branches is directly proportional to the expected number of substitutions that has occurred on that branch. The Melanesian branch (the most distant variants) appeared to have diverged a long time after the HTLV-II.

Subsequently, the Central African isolates diverged, followed by the West African. On this tree (Fig. 1b), the concordance between RFLP analysis and sequence clustering is again nearly perfect except for three isolates DIA, TSP1 and MELJ which are of RFLP subtype African (DIA) and Cosmopolitan (TSP1 and MELJ) respectively but clustered on sequence analysis within the Japanese cluster. The topology of this UPGMA tree indicated that both the Cosmopolitan and the Japanese subtypes are quite homogeneous and related to each other, whereas the two African subtypes appeared more divergent despite a common African origin.

By analysis of the LTR RFLP results on 180 HTLV-I proviral DNAs, covering the major HTLV-I endemic areas and including 75 new specimens, we demonstrated the existence of four main molecular subtypes which are mostly linked to the geographical origin of the infected individuals and not to the patient’s clinical status. This classification was validated by the LTR sequence comparative analysis which showed a nearly perfect correspondence between the molecular clusters based on three phylogenetic sequence analysis and the LTR.
subtypes based on RFLP profile. Furthermore, the phylogenetic analyses of 42 LTR HTLV-I sequences demonstrated that, within the African subtype, two main clusters existed; one corresponding to the West African samples and the other to Central African specimens. This allows us to propose a new classification of five major HTLV-I genotypes: Cosmopolitan, Japanese, West African, Central African and Melanesian.

Some of the specimens included in the present study have been analysed by sequencing of 522 base pair of the env gene (gp21) indicating similar geographical molecular clusters (Gessain et al., 1992, and unpublished data). Thus our proposal for a new classification of HTLV-I LTR subtypes also fits with the sequence analysis of the env regions, suggesting that, the HTLV-I genomic drift of both non-coding LTR region and coding env region evolved in parallel over centuries or millenia, despite the fact that the level of genetic variability is higher in the LTR than in the env gene (Komurian-Pradel et al., 1991). However, due to the lower genetic variability in the env gene, it is more difficult to segregate the Japanese type from the Cosmopolitan when using a small fragment of env gene for analysis than by LTR RFLP. Furthermore, the very simple RFLP technique can be used to rapidly and efficiently type any new HTLV-I ex vivo DNA even when available only in very small amounts, such as blood specimens collected on filter papers (Nerurkar et al., 1993).

Assuming similar mutation rates and selective forces among the different HTLV-I isolates the topology of the phylogenetic trees should reflect the duration of HTLV-I evolution and the migrations of virally infected populations in the recent and distant past. The results presented here allow us to put forward hypotheses concerning the routes of dissemination of this oncoretrovirus. The Cosmopolitan subtype is the most homogeneous HTLV-I molecular strain. Its dissemination could be recent, from a common ancestor towards multiple areas of the world, through recent migrations of HTLV-I-infected individuals including the slave trade from Africa to the Americas. (Gessain et al., 1992, 1994a, b). The five Iranian specimens (ATL, TSP/HAM or HTLV-I seropositive individuals), all from the town of Mashad, belong to the Cosmopolitan group, suggesting an introduction of the virus into this Middle East region over the past several centuries. This region, an ancient place of pilgrimage known to be a crossroad of population migrations in the past, has recently been shown to be highly endemic for HTLV-I and for associated diseases (Safai et al., 1992; Gabarre et al., 1993, R. Farid et al., unpublished data). Similarly, the Reunion Island strain, of Cosmopolitan type (confirmed by sequencing of gp21; Mahieux et al., 1994), suggests a recent introduction of HTLV-I in this area. This is in accordance with the known historical migrations demonstrating that this island has been populated during the last three centuries. Our findings do not support the hypothesis presented by others (Saksena et al., 1992) of a possible HTLV-I dissemination from the Indomalay region to the African continent through migrations in the Indian Ocean area.

Concerning the African subtype, the analysis of the LTR sequences of the 11 samples of this subtype demonstrated a greater sequence variability in this group than within the other subtypes (Cosmopolitan and Japanese). The Zairian strains appear so far to be the most distant African variants from the HTLV-I ATK prototype. Studies of the env genetic variability of a few samples from Central African countries demonstrated that there exist other molecular variants which are not restricted only to Zaire or Gabon (Fukusawa et al., 1987; Boeri et al., 1993; Gessain et al., 1994b and unpublished data). The diversity of these HTLV-I African strains may reflect a genetic drift during a long evolution in remote areas but also the probable occurrence of interspecies transmission between human and monkeys species infected by simian T cell leukaemia virus type I (STLV-I) (Saksena et al., 1993). Such a possibility has been recently suggested by the findings of a high degree of similarity (98%) between HTLV-I present in individuals living in the equatorial region of Zaire and STLV-I from chimpanzees (Koralnik et al., 1994).

The data based on the RFLP study of the Japanese specimens confirmed the existence of two different HTLV-I molecular subtypes in Japan which has been previously suggested on a small series (Komurian-Pradel et al., 1992). The first subtype (22%), belongs to the Cosmopolitan subtype, whereas the Japanese subtype represents the major (78%) subtype, being nearly exclusively restricted to Japan in our study. Similarly the LTR sequence analysis of 18 Japanese proviral DNAs showed the existence of two subtypes clustering apart on the phylogenetic trees. A recent study based on the analysis of nine Japanese isolates (including eight analysed in our present work) suggests that the presence of Cosmopolitan subtype both in Japan and in the Caribbean could be linked to the movement of Mongoloid people in the palaeolithic period from Asia to the Americas (Miura et al., 1994). However further genetic studies from a large variety of individuals from different Japanese areas and different Amerindian ethnic groups are necessary to confirm this interesting hypothesis (Ureta Vidal et al., 1994). In this context it is worthwhile to note that recent sequence data from some Indian HTLV-I isolates indicate the presence of both subtypes (Cosmopolitan and Japanese) in this region (Nerurkar et al., 1993, Hashimoto et al., 1993).

In spite of the fact that in the present study we
investigated only two samples from Melanesia, this appears sufficient to propose the existence of a Melanesian RFLP LTR subtype, which is molecularly quite distant from all other subtypes. The existence of such a distinct subset has recently been shown by genetic analysis of the env gene of other isolates from Papua New Guineans, Solomon Islanders and Australian Aboriginals (Sherman et al., 1992; Sakse et al., 1992; Gessain et al., 1991, 1993; Bastian et al., 1993). The presence of distant molecular variants of HTLV-I probably resulted from a genetic drift over several millennia in remote populations which originated 10000 to 40000 years ago from the Indomalay region (Yanagihara et al., 1991; Gessain et al., 1991, 1993; Sherman et al., 1992; Sakse et al., 1992).

Despite centuries or millennia of in vivo evolution, the very low level of mutation occurring in the transcrip-
tional enhancers of the 5' LTR of 12 new proviral DNA sequences demonstrates the existence of a high genetic constraint maintaining the critical role of these regulatory elements for viral expression. The consensus TGACC sequence contained within these enhancer regions (which responds to both cyclic AMP and HTLV-I Tax-induced proteins) was highly conserved in the 12 newly analysed specimens. The functional significance of the three observed changes at positions 102, 151, 255 in some of the Central African isolates is unknown.

Further molecular genetic studies from different isolated African and Amerindian populations, but also mongoloid populations from Asian regions including China, India (Nerurkar et al., 1993), Siberia (A. Gessain et al., unpubl. results), and Indian Ocean islands will be crucial to gain new insights on the origin and pathways of global dissemination of this human oncoretrovirus whose low degree of genetic drift in vivo may be useful as a marker for charting the migrations of infected human populations in the recent or distant past.

We are indebted to Drs A. Mahe, B. Rio, J. Gabarre, Y. Leguano, D. Moynet and O. Gaut for providing us with some of the specimens from Iran, Mali, Congo, Central African Republic, Gabon, French Guyana and Senegal. Furthermore we thank Dr Robert Bomford for helpful review of the manuscript. Financial support from the Agence Nationale de Recherches sur le SIDA (ANRS) is acknowledged, together with the World Laboratory (subproject MCD-2) and the Comité de Paris de la Ligue Nationale Contre le Cancer.

References


Short communication


Short communication


(Received 25 April 1994; Accepted 9 August 1994)