Distribution of Human T Cell Lymphotropic Virus Type 1 (HTLV-1) Subtypes in Brazil: Genetic Characterization of LTR and Tax Region

SIMONE KASHIMA,1 LUIZ CARLOS ALCANTARA,2 OSVALDO MASSAITI TAKAYANAGUI,3 MARCO AURELIO VALTAS CUNHA,4 BERNARDO GALVÃO CASTRO,2 MARIA SOCORRO POMBO-DE-OLIVEIRA,5 MARCO ANTONIO ZAGO,6 and DIMAS TADEU COVAS6

ABSTRACT

We report the molecular and epidemiological characterization of 128 human T cell lymphotropic virus type 1 (HTLV-1) isolates from Brazilian patients with different clinical manifestations of the infection. Thirty-two percent of the patients were asymptomatic, 44% had HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP), and 23% had adult T cell leukemia/lymphoma (ATLL). Phylogenetic analysis performed using part of the LTR region of the viral genome revealed that all Brazilian isolates belonged to theCosmopolitan subtype, with the following distribution within the Transcontinental subgroup: 81.6% within the Latin American cluster and 15.8% outside the Latin American cluster. Two isolates belonged to the Japanese subgroups. Molecular analysis of the tax region showed a high nucleotide similarity (~99%) with 41 prototype sequences, including the ATK-1 isolate. The mean number of nucleotide substitutions ranged from 1 to 8. Five specific nucleotide substitutions, C7401T, T7914C, C7920T, C7982T, and G8231A, were highly conserved among the Brazilian isolates (79.6%), with a frequency ranging from 81.6% to 100% in the sample group and from 18.4% to 24.1% in the prototypes used, suggesting the existence of a molecular signature. These changes were not correlated with a specific clinical status of the patients and could be a molecular characteristic of the HTLV-1 strains that circulate in Brazil.

INTRODUCTION

THE HUMAN T CELL LYMPHOTROPIC VIRUS TYPE I (HTLV-1) is the causative agent of a neurologic syndrome called HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP)1,2 and of the hematological disorder called adult T cell leukemia/lymphoma (ATLL).3,4 The HTLV-1 infection is endemic in Japan, the Caribbean, some regions of Central Africa, Oceania, and Middle East Mashadi Jews.5,6 South America is also considered to be an endemic area and HTLV-1 infections have been observed among Andean populations from Colombia, Ecuador, Peru, Bolivia, Argentina, and Chile.7-9

The genome of HTLV-1 is composed of approximately 9.0 kb distributed among structural (gag, env, and pol genes) and regulatory (pX and Rex regions) genes. Based on phylogenetic analyses of long terminal repeat (LTR) and env regions, seven different genetic subtypes of HTLV-1 have been proposed: (1) a, or Cosmopolitan, which is distributed worldwide10; (2) b, or Central African11; (3) c, a highly divergent Melanesian strain from Papua New Guinea and Australia12; (4) d, isolated from Central African Republic (CAR) pygmies and from two patients...
in Cameroon and Gabon\textsuperscript{13,14}; (5) e, isolated from a single sample from an Efe pygmy in the Democratic Republic of Congo (DRC); and (6) f, detected in an individual from Gabon.\textsuperscript{15} Recently, an HTLV-1 subtype g has been described.\textsuperscript{16} According to the LTR region, the most widespread and best-studied subtype, Cosmopolitan, is further divided into five subgroups: Transcontinental (A), Japanese (B), West African/Caribbean (C), North African (D), and Black Peruvian (E).\textsuperscript{10,17–19}

To date, a variable prevalence of HTLV-1 infections has been demonstrated in Brazil.\textsuperscript{20} However, up to now, few genotyping studies of HTLV-1-infected people in Brazil have been reported. Previous studies have reported that HTLV-1 strains from Brazil belong to the A subgroup of the Cosmopolitan subtype.\textsuperscript{21,22} In the present study, we report the molecular and epidemiological characterization of 128 HTLV-1 isolates from different geographic regions of Brazil. We also report, for the first time, the existence of a molecular signature in the tax region of HTLV-1 Brazilian isolates.

**MATERIALS AND METHODS**

**Study population, serology, and HTLV-1 subtyping**

The HTLV-1 patients studied here belong to the positive serology profile of blood donors of the Regional Blood Center of Ribeirão Preto and of patients from the Neurology Department of the University Hospital and the Instituto Nacional do Cancer, Rio de Janeiro. The study was approved by the Institutional Ethics Committee and all patients were informed about the procedures of sample collection and analysis and gave written consent to participate. Between December 1999 and May 2004, a 5-mm\textsuperscript{3} intravenous blood sample was collected from each patient into tubes containing EDTA anticoagulant. Serum/plasma samples from all donors were screened for HTLV-1 antibodies by enzyme-linked immunosorbent assay (ELISA) (rp21e-enhanced EIA Cambridge Biotech Corp., Worcester, MA).

**PCR**

DNA was extracted from the buffy coat (EIA-positive samples) using the SuperQuick DNA purification kit (ACGT, Piscataway, NJ). HTLV nested polymerase chain reaction (PCR) was performed in the LTR and tax regions to confirm the presence of proviral DNA and to differentiate between HTLV-1 and 2.

**LTR region.** First round PCR was performed on samples to amplify a fragment of the LTR gene using the primers LTR1 (5’ cattctactctacacggtctagctacatta 3’, position 8186–8221) and LTR2 (ccgaattcagggcacc 3’, position 8759–8740). A second round was carried out using nested primers LTRN1 (5’ tgaagaat acaccaacatccc 3’, position 8247–8266) and LTRN2 (5’ gaacgacgtcaccgctggtgat 3’, position 8724–8701). These primers amplify a fragment of 477 bp for the LTR region of HTLV-1, corresponding to nucleotides (nt) 8247–8274 of the ATK-1 isolate. Reactions were performed in 50-μl mixtures containing 500 ng of DNA, 1.0 U of Taq DNA polymerase, 50 mM KCl, 10 mM Tris–HCl, pH 8.3, 1.5 mM MgCl\textsubscript{2}, 0.8 mM each deoxynucleotide triphosphate (dNTPs), and 10 pmol of each primer. PCR cycling conditions consisted of one cycle at 94°C for 5 min and 35 cycles at 94°C for 90 sec, at 55°C for 90 sec, and at 72°C for 30 sec. Five microliters of the initially amplified product was used in the nested reaction. First and second rounds of amplifications were cycled as above.

**Tax region.** First round PCR was performed using the primers and conditions previously described by Furukawa et al.\textsuperscript{23} The amplified products were analyzed by 2% agarose gel electrophoresis followed by ethidium bromide staining.

**Nucleotide sequencing and phylogenetic analysis**

The tax and LTR region amplified products of 128 isolates were purified using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA) and directly sequenced with an automated ABI 377 Sequencer (Applied Biosystems Inc., Foster City, CA). Sequencing reactions were performed using the inner PCR primers and the ABI Prism Big Dye Terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA) according to the manufacturer’s instructions. The electropherograms were analyzed with the Sequencing Analysis software, version 3.3 (Applied Biosystems).

The multiple sequences of the HTLV-1 LTR and tax nucleotide sequences together with related HTLV-1 sequences from the GenBank/EMBL database were aligned using the Dambie program, version 4.0.24, and further edited with the GeneDoc program version 2.6. Neighbor joining (NJ) and maximum likelihood (ML) trees were constructed using the PAUP* 4.0b10 software (Sinauer Associates, Sunderland, MA). The F84 evolutionary model (which takes into account different substitution rates for transversions and transitions, as well as intersite substitution rate heterogeneity, as modeled using a γ-distribution) was selected as the best model for analysis of partial LTR sequences using Modeltest software version 3.7.\textsuperscript{25} The reliability of the NJ trees was assessed by analyzing 1000 bootstrap replicates. For ML trees, a heuristic search was performed with a subtree-pruning-regrafting branch swapping algorithm using the NJ tree as the starting material, including its optimized parameters. The likelihood ratio test (LRT) was used to calculate statistical support (expressed as p-values) for the branches. Trees were drawn with the TreeView software, version 1.4.\textsuperscript{25}

The tax protein region was analyzed with the Viral Epidemiology Signature Pattern analysis software (VESPA).\textsuperscript{26} For this purpose, we first compared the alignment of Brazilian strains (n = 84) with prototypes (n = 41). We then divided Brazilian tax sequences according to clinical status into (1) healthy asymptomatic carrier HAC, (2) HAM/TSP, and (3) ATLL, and proceeded with the comparison between healthy asymptomatic (HAC) × clinical disorders and HAM/TSP × ATLL.

**RESULTS**

**Epidemiological and clinical features of HTLV-1-infected patients**

We studied a total of 128 HTLV-1-positive samples from individuals belonging to distinct regions of Brazil (Fig. 1), most of them from the Southeast region (São Paulo and Rio de Janeiro).
States), corresponding to 77.8% of the isolates. Samples from the Northeast (13.2%, Bahia, Pernambuco, Paraíba, and Ceará states), South (0.8%), and Central West (0.8%) regions were also collected. Two samples (BRRJ028-98 and BRRJ127-93) were from Japanese immigrants.

All patients were evaluated for clinical status according to criteria previously described for ATLL and HAM/TSP. Of the individuals studied, 32% were HAC, 44% had HAM/TSP, and 23% had ATLL. None of the patients showed positive serology for other retroviruses, e.g., HIV-1 or HTLV-2. The sex, age, and ethnic background distribution of the blood donors are shown in Table 1. The HAM/TSP group showed a high frequency of infected females ($p < 0.05$), with a female to male ratio of 3.1:1. A high frequency of infection among African-Brazilian individuals was detected in both the ATLL and HAM/TSP groups (Table 1). Potential routes of transmission in this population were transfusion, sexual intercourse, intravenous drug use, and mother-to-child transmission.

![Brazil map showing the distribution of samples collected in this study.](image_url)

**FIG. 1.** Brazil map showing the distribution of samples collected in this study.

<table>
<thead>
<tr>
<th>Clinical status</th>
<th>Total (128)</th>
<th>Mean age (range)</th>
<th>Sex (%)</th>
<th>Ethnic background (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>ATLL</td>
<td>28</td>
<td>35.7 (14–71)</td>
<td>16 (57)</td>
<td>12 (43)</td>
</tr>
<tr>
<td>HAC</td>
<td>56</td>
<td>44.6 (19–72)</td>
<td>30 (54)</td>
<td>23 (42)</td>
</tr>
<tr>
<td>HAM/TSP</td>
<td>44</td>
<td>51.2 (32–74)</td>
<td>12 (28)</td>
<td>32 (72)$^{b}$</td>
</tr>
</tbody>
</table>

$^{a}$ATLL, adult T-cell leukemia/lymphoma; HAC, healthy asymptomatic carrier; HAM/TSP, tropical spastic paraparesis/human T leukemia virus type 1-associated myelopathy; ND, not determined.

$^{b}$p < 0.05 (Chi square test).
FIG. 2. Rooted NJ tree of eight HTLV-1 strains based upon a 420-bp fragment of the LTR region. The bootstrap values (above 50% and using 1000 bootstrap samples) on the branches represent the percentage of trees for which the sequences at one end of the branch form groups. Pyg19 and MEL 5 strains are used as outgroups. Geographical origin and ethnic origin are given in parentheses. Statistical evaluation of branch lengths results in some clades being better supported than others (**p < 0.001 or *p < 0.005). The Brazilian isolates are in bold. Symbols indicate the ethnic background of HTLV-1-infected individuals: (#) white people, (●) nonwhite (including mullatos and black people).
Phylogenetic analyses

The HTLV-1 sequences of Brazilian isolates were compared to the sequences deposited in the EMBL/GenBank database (98 prototypes for LTR and 41 prototypes for tax). We analyzed 420 bp and 1000 bp of the LTR and tax region, respectively. Phylogenetic trees were constructed and then evaluated by the NJ and ML methods. The reliability of the NJ topology was statistically evaluated using 1000 bootstrap replicates. The NJ tree (Fig. 2) shows that the majority of the new Brazilian isolates belong to the Transcontinental subgroup within the Latin American cluster (81.6%), while the other 15.8% are outside this cluster. As expected, the Japanese immigrants isolates clustered into the Japanese subgroup (1.6%). Additionally, we found two samples (BRRP438 and BRRJ275-95) that clustered within the South African isolates (South Africa cluster (1.6%). Additionally, we found two samples (BRRP438 and BRRJ275-95) that clustered within the South African isolates (TSP1, TBH1-3, RKl2Rrm, Afs911) with a high statistical support by ML ($p < 0.005$).

Eighty-four strains were selected according to the clinical status for molecular analysis of the tax gene, which was performed on the entire region of the gene (1.0-kb fragment), showing a nucleotide similarity of 99% with ATK-1 and 28 other reference sequences. Phylogenetic analysis showed that all strains from Brazil clustered with the tax A group, except sample BRRJ 127-97, which clustered with the tax B group (data not shown).

References ($n = 41$) and Brazilian tax sequences ($n = 84$) were aligned separately, and the two groups were then compared by analysis with the VESPA software. We found nucleotide differences of 1 to 8 positions. Analysis of the pattern of these substitutions showed five specific nucleotide substitutions at positions C7401T, T7914C, C7920T, C7982T, and G8231A, revealing the occurrence of a haplotype that was highly conserved among Brazilian isolates (79.6% of the samples). These substitutions occur at a frequency ranging from 81.6% to 100% in the group of Brazilian isolates, compared to 18.4–24.1% in the reference isolates (Table 2). In addition, we performed the analysis among the distinct groups: HAC ($n = 28$), HAM/TSP ($n = 33$), and ATLL ($n = 20$). We observed that 30.9% of HAM/TSP and 17.3% ATLL individuals presented the molecular signature pattern compared to 25.9% of the HAC group. No association was observed when we compared this signature pattern with the clinical status (Fisher’s exact test, $p$ value = 1.000).

### Table 2. Nucleotide Substitution Position in tax Region

<table>
<thead>
<tr>
<th>Position</th>
<th>7401</th>
<th>7914</th>
<th>7920</th>
<th>7982</th>
<th>8231</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATK-1</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Brazilian isolates ($n = 128$)</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>A</td>
</tr>
</tbody>
</table>

*Nucleotide positions were numbered according to the ATK-1 (accession number J02029).*

**DISCUSSION**

Studies of seroprevalence among Brazilian blood donors have shown variation in HTLV-1 prevalence ranging from 0.08% in the South to 1.80% in the North.28,29 A survey of blood donors between 1996 and 2004 (over 400,000 donations) showed that the HTLV-1/2 seroprevalence is about 0.2% in the Northeast region of São Paulo State, southeast of Brazil.

Considering the high level of miscegenation of the population and the geographic extension of the country, we evaluated the phylogenetic and molecular characteristics of HTLV-1 strains from different Brazilian states including the Northeast, Central West, and Southeast regions where different ranges of HTLV-1/2 frequencies are observed. Analysis of LTR nucleotide sequences showed that all isolates from Brazil belonged to the Transcontinental group of Cosmopolitan subtype A, except for two that clustered together with the Japanese subtype B, as shown by a highly significant maximum likelihood analysis (Fig. 2). In view of the large Japanese immigration at the beginning of the 1900s, it is not surprising to find this subtype in our country. We also identified two strains belonging to the South Africa cluster (1.6%) that had never been previously reported in Brazil. These isolates were isolated from white males from Ribeirão Preto (São Paulo state) and Rio de Janeiro (Rio de Janeiro State) cities, with no symptoms of HTLV-1-related diseases. These results showed that the majority (81.6%) of HTLV-1 subtypes circulating in the Southeast of Brazil clustered together with the Cosmopolitan type, Transcontinental subgroup, South American cluster. The similarity analysis showed that samples BA101, BA102, BRRP418, BRRP085, BRRP083, and BRRP425 were identical to the Bl2 from Peru, and therefore they clustered within the Latin American cluster. All of them presented HAM/TSP, excluding BA048, BA100, and BA101, which were not evaluated for clinical manifestations. Those samples were obtained from Afro-Brazilian populations, except BRRP083, BRRP338, and BRRP418, which belonged to white people. We did not identify any other African subtype among the samples studied; instead we had a large number of samples belonging to people of African-Brazilian ethnic background (approximately 45%). Considering these results as a whole, we can conclude that in Brazil the majority of the circulating HTLV-1 strains belong to the Cosmopolitan subtype. The analysis also showed that the majority of the Brazilian isolates formed a homogeneous group inside the Latin American cluster (Ia). However, some Brazilian sequences were scattered among the other clusters suggesting a variable degree of molecular diversity among circulating HTLV-1 strains.

In addition, we report for the first time the molecular analysis of the tax gene of HTLV-1 among Brazilian isolates, showing that these isolates harbor a molecular signature that distinguishes them from isolates from other geographic regions. These specific nucleotide substitutions were found at positions C7401T, T7914C, C7920T, C7982T, and G8231A, forming a haplotype that was observed in 79.6% of Brazilian isolates. No correlation was observed between these genetic polymorphisms of the tax region among Brazilian isolates and disease development. Although we found that changes in the tax region were not correlated with a specific clinical status of the patients, this
seems to be a geographic characteristic since molecular phylogeny analysis using the LTR region revealed that 98.3% of the HTLV-1 sequences clustered within the Cosmopolitan subtype, HTLV-1a.

In 2000, based on phylogenetic analysis, Furukawa et al. proposed that the HTLV-1 tax gene could be classified into two distinct subgroups, i.e., tax A and B, and that tax A probably was associated with the occurrence ofHAM/TSP. The phylogenetic analysis of the tax gene of Brazilian samples showed that all isolates but two clustered with the tax A group. Therefore, in the Brazilian isolates studied there was no association between tax A phylogeny and the clinical manifestations of HTLV-1 infection. These two isolates that grouped with tax B belonged to Japanese immigrants who developed chronic and acute forms of ATLL. As expected, these isolates also clustered among Brazilian isolates, we demonstrated by sequence analysis that infected patients harbored the HTLV-1a subtype and that different genomic regions of HTLV-1 should be analyzed in order to genetically characterize Brazilian strains. These studies will contribute to understanding the divergence of HTLV strains and the pathogenesis of this infection.

### SEQUENCE DATA

The GenBank accession numbers of the HTLV-1 LTR fragments included in the phylogenetic study are as follows. The other HTLV-1 reference strains were pyg19, L76310; Afs911, L72212; ITIS, Z32527; MT4LB, Z31661; MELS, L02534; HS35; D00294; GH78, D23639; CH26, D23690; Bo, U12804; OD, U12805; PY52, U12806; PY144, U12807; Ni1.Peru, Y16484; BI1.Peru, Y16481; Me1.Peru, Y16478; BI3.Peru, Y16483; BI2.Peru, Y16482; Qu1.Peru, Y16475; Me3.Peru, Y16480; Qu3.Peru, Y16477; Me2.Peru, Y16479 and N2.Peru, Y16487; RK14.Peru, AFO54627; ATM, J02030; ATK1, J02029; ATL-YS, U19999; BC12.1, U32557; BC11.2, U32552; H5, M57299; 7Rm, M811248; MT2, L03562; CR1, K02722; TBH1 to (4, L76026, L76025, L76034 and L76028; TBH6 to (7, L76030 and L76029; BOI, L36905; TSP1, M86840; RK12-Rum, AF3012; RK1-Iran, AF3010; AMA, CMC, FCR and MAQS, X88871-X88873 and X88876; KUW-1, L42253; Abl.A, U87264; CH, M69044; CAM, NAR, and NM1626, AFO63819–AF063821; KAPO, YAM, and TOTO, AFO76254–AF076256; CR015 and CR01, AFO76258 and AFO76259; NM1688; NIC, NM2082 and NM2015, AFO76261–AF076264; CR08, AFO76267. GenBank accession numbers of the HTLV-1 tax fragments: AFO45380 and AFO45381 (Spain), AFI33522 (Brazil), ABO45549, ABO45547, ABO45483, ABO45521, ABO45441, ABO45348, ABO45520, ABO45529, ABO45481, ABO45476, ABO45558, ABO45559, ABO45421, ABO36346, ABO36376, ABO36377, ABO36371, ABO45510, ABO45548, ABO45540, ABO36378, ABO36368, ABO36369, and J02029 (Japan), AF292000 (MT-2), HTU19949, AB181224 (Iran), S67443, SIAF139170, K30X, K34X, AFO40271 (Germany), AFO29264 (China), D13784 (Caribbean) and MEL-5.

The GenBank accession numbers of the HTLV-1 LTR and tax fragments sequenced in our laboratory and included in the phylogenetic analysis are as follows: LTR region sequences: DQ323750 to DQ323832. Tax region: DQ323833 to DQ323883.

### ACKNOWLEDGMENTS

This study was supported by the Center for Cell-Therapy (CTC)/Fundação Hemocentro de Ribeirão Preto-FUNDHERP and Fundação de Amparo à Pesquisa do Estado de São Paulo—FAESP. We also thank Rochele Azevedo and Elizabeth Mit-sue Hachiya for technical assistance.

### REFERENCES


Address reprint requests to:
Simone Kashima
Molecular Biology Laboratory
Regional Blood Center of Ribeirão Preto
Rua Tenente Catão Roxo, 2501
Ribeirão Preto 14051-140, Brazil
E-mail: skashima@hemocentro.fmrp.usp.br