HTLV-1 Infection in Blood Donors From the Western Brazilian Amazon Region: Seroprevalence and Molecular Study of Viral Isolates

Aline Cristina Mota-Miranda, Sérgio Pereira Araújo, Juarez Pereira Dias, Denise Duizit Colin, Simone Kashima, Dimas Tadeu Covas, José Tavares-Neto, Bernardo Galvão-Castro, and Luiz Carlos Junior Alcantara

1Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil
2Universidade Federal da Bahia, Salvador, Bahia, Brazil
3Fundação Hospital Estadual o Acre, Rio Branco, Acre, Brazil
4Hemocentro de Ribeirão Preto, São Paulo, Brazil
5Escola Bahiana de Medicina e Saúde Pública, Fundação Bahiana para o Desenvolvimento das Ciências, Salvador, Bahia, Brazil

To determine the seroprevalence of HTLV-1 in Brazil, and to review the virus molecular epidemiology in this Amazon population (Rio Branco-Acre), 219 blood donors were screened for HTLV-1. Only one case of infection (0.46% seroprevalence) was detected during July 2004 screening at the Acre Hospital Foundation (FUNDACRE). Neighbor-joining and Maximum Likelihood phylogenetic analyses of two (n = 2) complete LTR region sequences were performed with the PAUP* software. Since the HTLV-1 envelope surface (gp46) and transmembrane (gp21) glycoproteins are important for virus fitness, three envelope glycoproteins sequences (n = 3) were analyzed using the Prosite tool to determine potential protein sites. Phylogenetic analysis demonstrated that the new isolate described in this study, and the unpublished LTR strain described in a previous report belong to the Transcontinental subgroup of the Cosmopolitan subtype, inside the Latin American cluster. A similar result was obtained when submitting, to the Automated Genotyping System, three LTR partial sequences from a previous study of the seroprevalence of HTLV-1 in the same Amazon population. In all analyzed env sequences, the potential protein site was found: two PKC phosphorylation sites at amino acid (aa) positions 310–312 and 342–344, one CK2 phosphorylation site at 194–197aa, three N-glycosylation sites at 222–225aa, 244–247aa and 272–275aa, and a single N-myristylation site at 327–338aa. In conclusion, potential protein sites described in HTLV-1 gp46 and gp21 confirm the presence of conserved sites in the HTLV-1 envelope proteins, likewise phylogenetic analysis suggests a possible recent introduction of the virus into North Brazil. J. Med. Virol. 80:1966–1971, 2008. © 2008 Wiley-Liss, Inc.

KEY WORDS: LTR region; glycoproteins; phylogenetic analysis; protein sites

INTRODUCTION

Human T-cell lymphotropic virus type 1 (HTLV-1) is a retrovirus spread widely throughout the world and is related to adult T-cell leukemia lymphoma (ATLL) [Poiesz et al., 1980] and tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM) [Gessain et al., 1985; Osame et al., 1986]. This infection has prevalence rates of more than 30% in Southern Japan, in the Caribbean Basin and some African regions, and to a lesser extent in Latin America [Proietti et al., 2005]. HTLV-1 exists in much of South America, including Brazil, Argentina, French Guyana, Chile and Colombia [Galvão-Castro et al., 1997; Yamashita et al., 1999; Dourado et al., 2003]. In Brazil, whose population is a mixture of Amerindians, Africans and Europeans, 2.5 million people are infected with HTLV-1.
According to epidemiological data, the prevalence of HTLV-1 in the southeast of Brazil (São Paulo) and the north (Manaus) regions are about 0.18–0.41% and 0.08%, respectively [Carneiro-Proietti et al., 2002], while the city of Salvador (Northeast region) has the highest HTLV-1 prevalence in Brazil [Galvão-Castro et al., 1997; Dourado et al., 2003; Mota et al., 2006]. In the Amazon region (Rio Branco-Acre), a study has estimated that the seroprevalence rates of HTLV-1 and HTLV-2 infection are about 0.08% and 0.03%, respectively [Colin et al., 2003]. In addition, in that geographic region, HTLV-1 has been found among urban and rural populations, while HTLV-2 is found mainly in the Amerindian people [Ishak et al., 1995; Vallinoto et al., 1998; Ishak et al., 2003].

According to the analysis of HTLV-1 LTR region, six genetic subtypes have been suggested [Miura et al., 1994]: a or Cosmopolitan (worldwide distribution); b or Central African; c or Melanesian; d, from Central African pygmies, Cameroon and Gabon; e, from the Democratic Republic of Congo; f, isolated initially from an infected individual from the Democratic Republic of Congo, and also described in Cameroon; and g, described recently as a new subtype in Cameroon. The Cosmopolitan subtype is divided in five different subgroups: A-Transcontinental, B-Japanese, C-West African/Caribbean, D-North African, and E-Black Peruvian. Up to now several molecular epidemiology studies has revealed the vast majority of the Brazilian HTLV-1aA strains could have been originated from several African lineages [Van Dooren et al., 1998; Yamashita et al., 1999; Alcantara et al., 2003].

The genetic contribution of the virus proteins is related to their important role in viral fitness. The surface glycoprotein (gp46) subunit is involved in cellular receptor recognition, while the transmembrane glycoprotein (gp21) subunit anchors gp46 to the cell and plays a major role in the post-binding steps of the fusion process. In this context, the viral glycoproteins are the first target of the immune system, and the posttranslational modifications of the virus proteins are essential for HTLV-1 fitness, assembly and immune escape, as they are for HIV infection [Nagy et al., 1983; Palker et al., 1989; Reitter et al., 1998].

In the present study, the seroprevalence of HTLV-1 was determined in blood donors from Acre. Since the distribution of the virus is considered to be related to the anthropological background and past human movements, a phylogenetic characterization of the viral isolates was also carried out. In addition, the analysis of envelope proteins could provide insight on the importance of the viral envelope protein in infection and immune escape.

**METHODS**

**Study Subjects and Serology**

Blood samples (219) were collected from blood donors between July 2004 and July 2005 at the Acre Hospital Foundation (Fundação Hemocentro de RioBranco-Acre), and processed at the Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Muniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil. Serum samples were screened for antibodies against HTLV-1 and 2: ELISA (HTLV, enhanced, EIA, Cambridge Biotech Corporation, Worcester, MA) and Western blot (HTLV Blot 2.4, Genelabs Diagnostics, Science Park Drive, Singapore). DNA was extracted using the GFX genomic blood DNA purification kit (Amersham Pharmacia Biotech, Piscataway, NJ).

**Phylogenetic Analysis**

Nested PCR was performed on the unique HTLV-1 positive sample, identified in this serologic screening, and on the four positive samples from a previous serologic characterization study [Colin et al., 2003]. LTR region was amplified as two overlapping fragments: a 5′ LTR-gag segment of 473 bp and a tax-3′ LTR segment of 479 bp, as previously described [Alcantara et al., 2006]. The *env* gene were also amplified in three positive samples from a previous serologic characterization study [Colin et al., 2003], using methods published previously [Yang et al., 1997]. The PCR products were purified from a 1% agarose gel electrophoresis using the Qiaquick Gel Extraction kit (Qiagen, Hilden, Germany) and sequenced in an ABI Prism 3100 DNA Sequencer using Taq FS Dye terminator cycle sequencing. The same inner PCR primers were used in the sequencing reactions.

The phylogenetic analysis, based on the complete LTR region, included reference strains from different geographic regions and distinct ethnic groups corresponding to all subtypes and described subgroups were selected from the NCBI/Nucleotide Sequence Database (GenBank). To better understand the epidemiology of HTLV-1 in the Amazon region and to obtain more genetic information about virus isolates from this region, four sequences (one complete and three partial) generated in a past study [Colin et al., 2003] were also genotyped by the phylogenetic analysis performed in this study. These sequences were also submitted to GenBank after analysis. These sequences were aligned using the Clustal X software [Jeanmougin et al., 1998] and edited manually using the GeneDoc program [Nicholas et al., 1997]. The Tamura Nei (TrN + G) evolutionary model (which takes into account different substitution rates for transversions and transitions, as well as inter-site substitution rate heterogeneity, using a γ-distribution) was selected using the Modeltest software [Posada and Krandall, 1998] as the best model. The neighbor-joining (NJ) and maximum-likelihood (ML) trees were generated using the PAUP* 4.0b10 software [Swofford, 1998]. The NJ tree was constructed with an optimized nucleotide substitution rate matrix and a γ-shape parameter (alpha parameter = 0.811083). The reliability of the NJ trees was assessed by analyzing 1,000 bootstrap replicates. For the ML tree, an 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 (RT) method was used to calculate statistical support for the branches: \( P < 0.001 \) (highly significant * * *) and \( P < 0.005 \) (significant *). Bootstrap and ML supports were added to NJ tree that was drawn with TreeView 1.4 software [Page, 1996]. The genotyping of three partial 3′LTR sequences was performed using the LASP HTLV-1 Automated Genotyping Tool (http://lasp.cpqgm.fiocruz.br).

**Potential Protein Site Analysis**

The genetic information obtained from the proteins was possible through the protein potential sites using the GeneDoc software [Nicholas et al., 1997] and the Prosite tool, as described previously [Mota-Miranda et al., 2007]. The protein sequences included in these analyses were generated during a previous seroprevalence study [Colin et al., 2003], and all the results were also compared with other Brazilian reference sequences.

**HTLV-1 Intra-Country Diversity**

The genetic distances were measured, within and between four distinct groups formed with LTR strains originated from Brazil (Salvador, Fortaleza, Acre) and Peru. The Kimura 2-\( \alpha \)-parameter model was used with a distance matrix implemented in the MEGA 3.0 package [Kumar et al., 1994], and the standard error computation was obtained by Bootstrap analysis (1,000 replicates).

**Nucleotide Sequence Accession Numbers**

Any of the eight sequences included in this study was analyzed or published previously. The nucleotide sequence accession numbers of the new HTLV-1 env and LTR sequences in this study are: env - ENVAC57 (EU392161), ENVAC69 (EU392162), ENVAC204 (EU392163); complete LTR- AC181 (EU392159), AC042 (EU392160); partial LTR- LTR5AC069 (EU392164), LTR5AC129 (EU392165), LTR5AC174 (EU392166).

**RESULTS**

The overall prevalence of HTLV-1 among blood donors from Rio Branco-Acre was 0.46% (1/219) and the phylogenetic analysis of the entire LTR region of this single new sequence (AC181) showed that it belongs to the Transcontinental subgroup of the Cosmopolitan subtype (Fig. 1) with a 71% bootstrap value (\( P < 0.001 \) for ML). This finding is quite similar when comparing this new sequence (AC181) with another (AC042) from a previous seroprevalence study [Colin et al., 2003]. The other three sequences (AC069, AC129, and AC174), corresponding to the partial 3′LTR region (350 pb), submitted to the Automated Genotyping System, were classified as a Transcontinental (A) subgroup of the Cosmopolitan (a) subtype.

The three env gene sequences, corresponding to the gp46 N-terminal region (179–320aa) and the gp21 C-terminal region (1–43aa) were submitted to the GeneDoc software using the Prosite tool to determine potential protein sites. Three different types of post-translational modifications were found into the gp46 protein: one PKC phosphorylation site at 310–312 amino acids (aa); one CK2 phosphorylation site at 194–197aa; and three N-glycosylation sites at 222–225aa, 244–247aa and 272–275aa. Otherwise, two different posttranslational modifications were identified into the gp21 protein: one N-myristylation site at 7–18aa and one PKC phosphorylation site at 22–24aa.

**DISCUSSION**

The HTLV-1 infection seroprevalence described in this study is higher than that reported previously in blood donors from Rio Branco-Acre (0.11%), between 1998 and 2001 [Colin et al., 2003]. Nevertheless, it is important to note that this new study has a small group population, and so the seroprevalence could be less if a larger number of individuals were screened.

To our knowledge, this is the first study considering the molecular epidemiology and protein sites of HTLV-1 isolates from the Western Amazon population of Rio Branco-Acre. Previous studies have demonstrated [Colin et al., 2003] that the frequency of phenotypic characteristics from the Amerindian mixed racial group in Acre’s population is 33.3%. Epidemiological data suggest that routes of transmission and risk factors associated with infection could be different among distinct places; therefore it is important to study the molecular profiles of viral isolates in different populations in Brazil.

The phylogenetic analysis performed in this study has identified two isolates from the Transcontinental (A) subgroup of the Cosmopolitan (a) subtype. This finding is in agreement with others studies that have indicated the extensive presence of this subgroup in viral isolates from Brazil [Alcantara et al., 2003]. The contribution of genetic studies indicates at least two hypotheses that give some support to the origin of this subgroup of the HTLV-1 Cosmopolitan Subtype in the New World. The first hypothesis suggests that the introduction of
Fig. 1. Rooted neighbor-joining tree of Brazilian HTLV-1 LTR region (615bp) isolates. HTLV-1 subtype reference sequences were obtained from the GenBank/EMBL databases. The bootstrap values >50% were considered and added in the tree. Mel5 was used as outgroup. Geographic origin is given between parentheses. The new LTR sequences included in this analysis are in bold. ** and * means that the ML method was highly significant ($P < 0.01$) or significant ($P < 0.05$), respectively.
the virus occurred during the first Paleo-Indian migration across the Bering Straits around 40,000–10,000 years ago [Miura et al., 1994]. The second hypothesis suggests that the virus was imported by the post-Columbian African slave trade, around 400 years ago [Van Dooren et al., 1998].

The fact that the HTLV-1 isolates from Rio Branco have grouped into two separate and well-defined clades, and that these two clades demonstrate small divergence, could suggest a recent and separate spread of the HTLV-1a virus in this Latin American population. A similar conclusion was considered in a recent Peruvian HTLV-1 molecular study [Zehender et al., 2007], when 3 of 10 new HTLV-1 isolates segregated within the main South African cluster, while seven clustered into the two known Latin American clusters.

The higher genetic diversity among Acre strains (2.7%) suggests a previous migration of the virus, across the Bering Strait to the American Continent, when comparing the lower genetic diversity among Northern Brazil strains (Salvador (1.2%) and Fortaleza (0.7%)). Comparing the diversity among HTLV-1 isolates, it was possible to identify a lower distance between Acre and Fortaleza strains. This suggests that the human migration during the Second World War from Fortaleza to the Amazon region, and back from Amazon region to Fortaleza, could explain a recent introduction of the virus in this population. The difference between genetic diversity calculated among Acre and Salvador isolates can also give support to the hypothesis that the slaves in Salvador were traded mainly from the West Coast of Africa, while those in the Amazon or North of Brazil came from the East Coast of Africa. Comparing the genetic diversity among Amazon region and Peruvian strains (2.7% × 2.1%), the relative relationship between these isolates could confirm, as postulated previously, that the genetic admixture of the black Latin Americans with other Latin American ethnic groups, such as Quechuas and Mestizos, giving rise, probably, to HTLV-1a in these indigenous Latin American populations [Van Dooren et al., 1998].

In the present report, four different posttranslational modifications were identified for the gp46 and gp21 glycoproteins, with highly conserved characteristics. It has been postulated that these phosphorylation, N-glycosylation and myristylation sites are important for the production of functional virus particles. The myristylation site, in the gp21 protein, is commonly associated with viral assembly [Bouamir et al., 2003]. The glycosylation sites are associated with virus protection from the immune system [Reitter et al., 1998], since these posttranslational modifications could prevent the recognition of infected cells by antibodies.

The potential protein sites identified in the Acre sequences are exactly the same, as those described previously in other Brazilian sequences [Mota-Miranda et al., 2007; Queiroz et al., 2007]. The presence of these conserved sites in the Env amino acid sequences suggests that the virus has maintained them as a strategy for escaping the immune system.
HTLV-I in the Western Brazilian Amazon Region


Queiroz AT, Mota-Miranda AC, de Oliveira T, Moreau DR, Urpia Cde C, Carvalho CM, Galvão-Castro B, Alcântara LC. 2007. Re-mapping the molecular features of the human immunodeficiency virus type I and human T-cell lymphotropic virus type 1 Brazilian sequences using a bioinformatics unit established in Salvador, Bahia, Brazil, to give support to the viral epidemiology studies. Mem Inst Oswaldo Cruz 102:133–139.


