

# Molecular Characterization of the Human T Cell Lymphotropic Virus Type 2 Long Terminal Repeat Region: A Discussion about Possible Influences at Viral Gene Expression

Fernanda K. Barreto,<sup>1</sup> Felipe F.A. Rego,<sup>1</sup> Loiana M. Fonseca,<sup>1</sup> Bernardo Galvão-Castro-Filho,<sup>2</sup> Thessika H.A. Araújo,<sup>1</sup> Aline Cristina A. Mota-Miranda,<sup>2,3</sup> Joana P. Monteiro-Cunha,<sup>2,3</sup> and Luiz Carlos J. Alcantara<sup>1</sup>

## Abstract

This study aimed to identify nucleotide signatures in the promoter region of human T cell lymphotropic virus type 2 (HTLV-2) isolated from infected individuals from Salvador, Brazil and in sequences from the GenBank database. DNA samples from HTLV-2-infected individuals were submitted to nested polymerase chain reaction (PCR) and sequencing, and molecular analyses were performed using bioinformatics tools. The phylogeny of HTLV-2 strains isolated from patients from Salvador reveals that all sequences were subtype c. One hundred and fifty-one sequences from GenBank were selected, among which 30 belong to subtype a, 88 to subtype b, 32 to subtype c, and one to subtype d. Subtype-specific signatures were identified as well as mutations resulting in loss or gain of motifs important to transcription regulation. The subtypes a and b have two E box motifs, while subtypes c and d have only one. These polymorphisms may impact viral fitness and infection outcome and should be more closely investigated.

**H**UMAN T CELL LYMPHOTROPIC virus type 2 (HTLV-2) was described in 1982.<sup>1</sup> Currently this virus is found mainly among Brazilian Amerindians, African populations, and intravenous drug users (IVDU) from the United States, Europe, and Asia.<sup>2</sup> HTLV-2 is not clearly identified as the etiologic agent of human pathologies, although it seems to be linked to neurological disorders and appears to be associated with an increased incidence of autoimmune diseases and respiratory tract infections.<sup>3,4</sup>

Viral and host factors are probably involved in determining the infection outcome. It is believed that small variations in the HTLV-2 long terminal repeat (LTR) region may modify the binding ability of transcription factors and this could influence gene expression.

Although most of the HTLV-2 genome remains stable, considerable variations are noted in the promoter region of this retrovirus. Phylogenetic analysis, using this region and the *env* gene, demonstrated three main subtypes of HTLV-2: HTLV-2a, HTLV-2b,<sup>5</sup> and HTLV-2d.<sup>6</sup> In addition, one cluster

within subtype a, formed exclusively by viral isolates from Brazil, was described in 1996, and it is called HTLV-2c.<sup>7</sup> Despite this genetic variability, to date there are no reports describing nucleotide signatures specific to the viral subtypes or evaluating the consequences of nucleotide variations at the gene expression.

Therefore, this study was performed with the objective of identifying nucleotide variations (signatures) in the HTLV-2 promoter region of isolates from different geographic regions and from different viral subtypes.

Blood samples were collected from 15 infected individuals followed at the HTLV reference center of Bahia School of Medicine and Public Health, located in Salvador city, Northeastern Brazil. All patients were monoinfected with HTLV-2. The Ethnic Committee of Centro de Pesquisa Gonçalo Moniz/FIOCRUZ approved this study. Informed consent was obtained from all patients.

DNA was extracted using a spin column QIAamp DNA Blood Minikit (Qiagen) following the manufacturer's

<sup>1</sup>Laboratório Hematologia, Genética e Biologia Computacional, Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil.

<sup>2</sup>Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia, Brazil.

<sup>3</sup>Universidade Federal da Bahia, Instituto de Ciências da Saúde, Salvador, Bahia, Brazil.

protocol. Samples were submitted to a nested polymerase chain reaction (PCR) with outer primers BSQF6/BSDR3 and inner primers BSQF2/BSDR4, to amplify a 672-pb fragment from the HTLV-2 LTR region.<sup>8</sup> All amplified products were submitted to an electrophoresis assay, using a 1% agarose gel with Syber Safe DNA (Invitrogen).

The PCR products were purified using the PureLink PCR Purification Kit (Invitrogen) and sequenced using the ABI Prism 3100 DNA Sequencer (Applied Biosystems Inc., Foster City, CA) with the Taq FS Dye terminator (Applied Biosystems) cycle sequencing. The same inner primers BSQF2/BSDR4 were used in the sequencing reactions and the electropherogram was analyzed with SeqScape software (Applied Biosystems SeqScape Software v.2.5). In parallel, we performed data mining for HTLV-2 LTR region sequences deposited in the GenBank database and selected 151 sequences, which were longer than 650 pb. These sequences are from different geographic regions and belong to different populations.

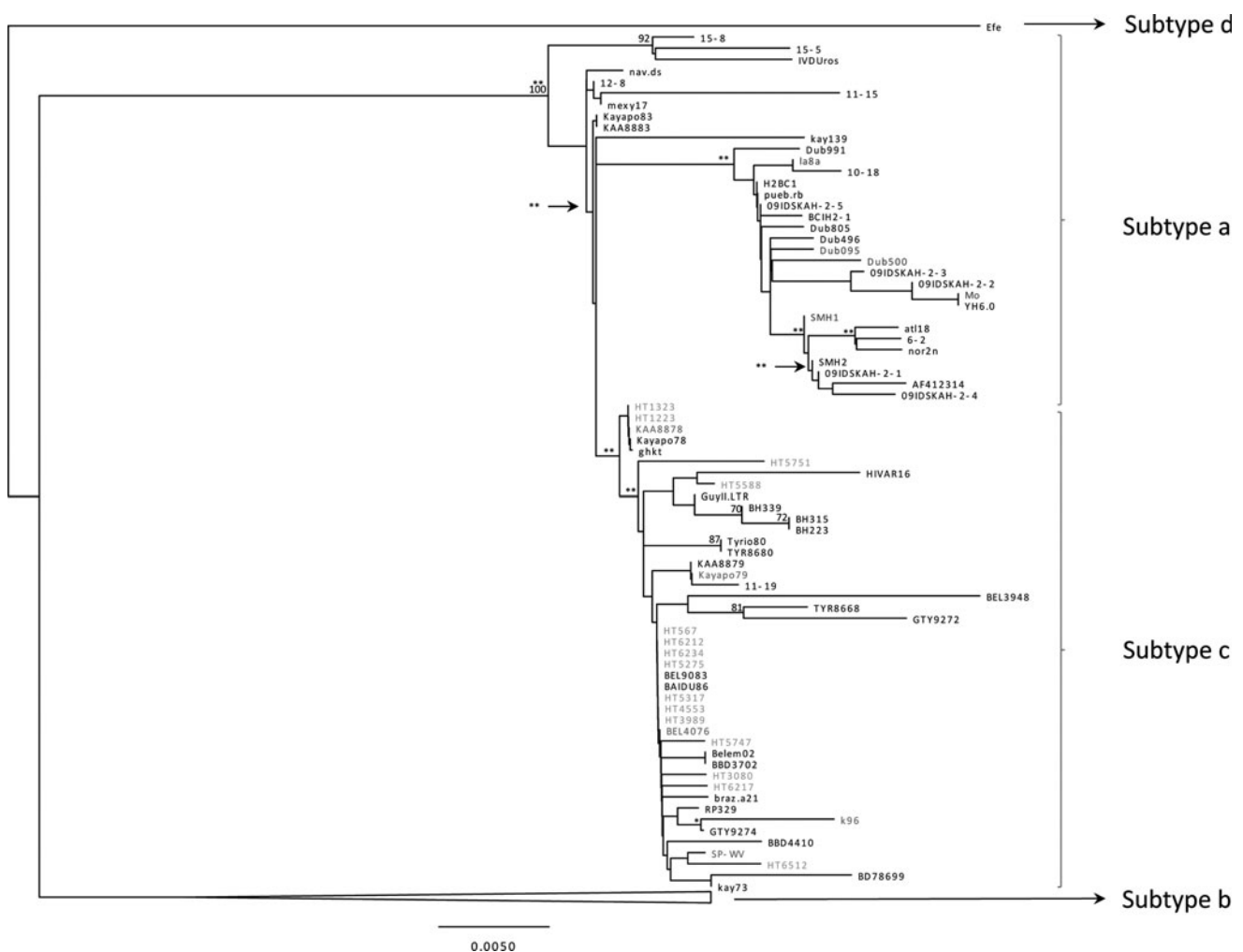
Clustal X software was used to perform the multiple sequence alignment with the new generated sequences and the sequences recruited from the GenBank database.<sup>9</sup> Finally, the multiple alignment was manually edited in the BioEdit program.<sup>10</sup>

Neighbor-joining (NJ) and maximum-likelihood (ML) trees were obtained with the PAUP 4.0 program using the HKY85 nucleotide substitution model.<sup>11</sup> The reliability of NJ trees was evaluated by analyzing 1,000 bootstrap replicates. The trees were visualized with the FigTree v.1.3.1 program.

The sequences were organized into groups with the same subtype, and, using the BioEdit program, a consensus sequence was created to each set of subtype sequences. The search for motifs and different molecular characteristics among the sequences from different subtypes was performed using the Geneious v.5.6.5 program.<sup>12</sup> The genetic distances were measured within and between the different subtypes using the MEGA 5.05 program.<sup>13</sup>

Currently the GenBank database provides 347 HTLV-2 LTR sequences. From this set of sequences, 151 were longer than 650 pb and, therefore, were included in the study. Phylogenetic analysis was performed using these 151 sequences and the new 15 HTLV-2 LTR sequences (672 pb) generated in this study.

Our analyses showed that 30 of the 151 sequences were subtype a, 88 sequences were subtype b, 32 were subtype c, and 1 was subtype d. All new generated HTLV-2 LTR sequences were subtype c (Fig. 1). The NJ phylogenetic analysis has ML support.



**FIG. 1.** Phylogenetic analysis of the long terminal repeat (LTR) region using a neighbor-joining (NJ) approach. The Efe strain was used as an outgroup. The new samples are shown in black and the sequences from GenBank are shown in gray. Statistical evaluation of branch lengths results is shown as  $**p < 0.01$  and  $*p < 0.05$ .

TABLE 1. GENETIC DISTANCES IN DIFFERENT SUBTYPES OF HTLV-2 USING THE MEGA 5.05 PROGRAM

	Subtype a	Subtype b	Subtype c	Subtype d
Subtype a	0.009	0.049	0.014	0.072
Subtype b	0.049	0.008	0.047	0.052
Subtype c	0.014	0.047	0.007	0.072
Subtype d	0.072	0.052	0.072	#

(#) The subtype d corresponds to only one sequence and therefore it is not possible to calculate the genetic distance into this subtype.

The 166 sequences were submitted to molecular characterization through the comparative analysis of consensus sequences. It was possible to identify seven nucleotide variations, called signatures, in the subtype a isolates, among which four (14T, 116T, 117T, and 741T) were present in all 30 sequences. Twelve signatures were identified among the subtype b strains, and six of them (114A, 166A, 271A, 275G, 629T, and 661C) were present in 100% of the b sequences. Analyzing the subtype c sequences, we identified only one signature (401T), which was found in 43 sequences. It is important to note that some signatures of subtype a are also frequently found in subtype c. The genetic distances measure demonstrated that genetic distance between sequences from subtypes a and c was lower than the distance between the other subtypes, and the genetic distance between sequences from subtypes a and d was the same as the calculated distance between subtypes c and d (Table 1). This probably occurs because HTLV-2c is considered a subcluster of HTLV-2a, since it has a similar Tax protein to HTLV-2b and the *env* gene and an LTR region similar to the HTLV-2a subtype.<sup>6</sup>

Using the consensus sequences obtained from each set of specific subtype sequences, the search for transcription factors was performed. Some motifs for transcription factors described previously in eukaryotic cells were sought in these

TABLE 2. TRANSCRIPTION FACTOR MOTIF IDENTIFICATION AMONG THE SUBTYPE CONSENSUS SEQUENCE OF THE HUMAN CELL LYMPHOTROPIC VIRUS TYPE 2 LONG TERMINAL REPEAT REGION

Transcription factor motif	Nucleotide position <sup>a</sup>	Subtype consensus sequences
CRE	98–102	a, b, c, d
SRE	109–118	b, c
ETS	119–122	a, b, c
CRE	138–142	a, b, c, d
ETS	156–159/162–165	a, b, c, d
Poly(A)	180–185	b, d
CRE	231–235	a, b, c, d
Poly(A)	269–274	a, b, c, d
ETS	275–278	b
TATA	287–292	a, b, c, d
ETS	361–364	a, b, c, d
E box	738–743	a, b
ETS	752–755	a, b, c, d
E box	762–767	a, b, c, d
ETS	780–783	a, b, c, d

<sup>a</sup>The nucleotide position was done using the reference sequence Mo (M10060).

sequences: TATA box, CAT box, CG box, E box, NF- $\kappa$ B, CRE, GATA box, ETS, SRE, SP1, poly(A), and CpG islands. It was not possible to identify many differences regarding the transcription factor binding sites appearance and frequency between the consensus sequences, and it was also not possible to identify all the above described transcription factor motifs in the consensus sequences (Table 2). The CpG islands are located at the same site (nucleotide position: 293) in all subtypes (Fig. 2). The nucleotide sites of all molecular analyses were identified using the reference sequence Mo (M10060).

Considering that the transcription factor motifs are important in regulating the transcription of HTLV-2 provirus, mutations associated with the loss or gain of any of these motifs could be responsible for changes in viral gene expression. Studies have shown that mutations in the E box region, for example, may increase or decrease transcription of viral genes, depending on the cell type in which the mutation occurred. Studies using the bovine leukemia virus (BLV) and performing site-directed mutagenesis assays suggest that mutations that result in the loss of the E box motif increase the expression of the LTR region and therefore these motif could be associated with transcription repression.<sup>14</sup>

In this study, we did not find geographic signatures or signatures among intravenous drug users compared with infected tribes and African populations. However, we have identified subtype-specific signatures, comparing the 15 new sequences and the 151 already published sequences. We have also demonstrated that the 15 new HTLV-2 isolates from Salvador belong to subtype c. Moreover, we have found mutations that result in loss or gain of the motifs important in regulating viral gene transcription. Therefore we believe that these results should be more closely investigated, because the regulation of gene transcription can influence the infection outcome.

## Sequence Data

The GenBank accession numbers for the sequences included in the phylogenetic analysis are as follows: Dub095 (AF032989), Dub500 (AF032990), nav.ds (U10257), 15-8 (U73015), 15-5 (U73020), IVDUros (AF054272), 12-8 (U73008), mexy17 (L42510), 11-15 (U73021), Dub805 (AF175467), Dub991 (AF032993), 10-18 (U73019), la8a (U10256), H2BC1 (AF115495), BCIH2-1 (AF185282), pueb.rb (U10262), 09IDSKAH-2-5 (JN247462), Dub496 (AF032992), 09IDSKAH-2-4 (JN247461), 09IDSKAH-2-3 (JN247460), 09IDSKAH-2-2 (JN247459), YH6.0 (NC001488), SMH1 (Y09147), SMH2 (Y09148), 09IDSKAH-2-1 (JN247458), AF412314, atl18 (U10252), nor2n (U10258), 6-2 (U73022), Mo (M10060), Kayapo83 (AF139390), KAA8883 (AF306733), ghkt (L42507), kay139 (L42508), Tyrio80 (AF139391), TYR8680 (AF306730), Kayapo78 (AF139388), KAA8878 (AF306731), Kayapo79 (AF139389), 11-19 (U73013), KAA8879 (AF306732), k96 (AF326584), RP329 (AF326583), HIVAR16 (JN222964), GuyILTR (AF262408), BH339 (AY509602), BH223 (AY509600), BH315 (AY509601), BEL9083 (AF306726), BAIDU86 (AF401492), BBD3702 (AF306735), Belem02 (AF139392), SP-WV (AF139382), TYR8668 (AF306729), GTY9272 (AF306727), GTY9274 (AF306728), BEL3948 (AF306724), braz.a21 (U10253), BBD4410 (AF306734), BEL4076 (AF306725), BD78699 (DQ235706), kay73 (L42509), Efe (Y14365), G2 (AF074965), G12 (L11456), wyu1 (HTU12792), BDAR6 (JN222947), 4-10 (HTU73016), 6-4 (HTU73018), 12-7 (HTU73012), pueb.ag (HTU10261), sem1051 (HTU10264),





Pyl36 (GQ330488), Pyl115 (GQ330493), 330110 (GU212849), 230102 (GU212848), penn7a (HTU10260), 2-4 (HTU73011), 15-21 (HTU73017), HIVAR4 (JN222952), HIVAR3 (JN222951), HIVAR12 (JN222960), HIVAR5 (JN222953), wyu2 (HTU12794), Pyl39 (GQ330490), HIVAR6 (JN222954), Pyl38 (GQ330489), Pyl41 (GQ330491), Pyl42 (GQ330492), Pyl116 (GQ330494), Pyl124 (GQ330496), PYGCAM-1 (Z46888), pilaga (AF054271), BDAR2 (JN222943), BDAR4 (JN222945), 14-17 (HTU73009), NRA-P (L20734), FOR6 (AF054273), Pyl121 (GQ330495), HIVAR10 (JN222958), sem1050 (HTU10263), HTLV-II-Gab (Y13051), BDAR3 (JN222944), Gab1080FC (EU444100), 550602 (GU212854), 880701 (GU212850), 670802 (GU212851), 500101 (GU212852), 590101 (GU212853), ita47a (HTU10254), 324 (L77243), RVP (L77244), JÁ (L77239), DP (L77237), ny185 (HTU10259), 9-17 (HTU73014), JAN (L77241), JL (L77240), SP21 (AY442384), ita50a (HTU10255), I-AM (Y09149), I-OG (Y09154), I-T (Y09151), I-OV (Y09155), SP17 (AY442380), SP12 (AY442375), SP23 (AY442386), SP22 (AY442385), SP20 (AY442383), SP6 (AY442369), SP8 (AY442371), span130 (HTU10266), BF (L77236), RC (L77235), AA (L77238), SP2 (AY442365), 130 (L77242), SP4 (AY442367), SP15 (AY442378), SP7 (AY442370), SP26 (AY442389), SP14 (AY442377), SP16 (AY442379), PortNn (AY622978), SP19 (AY442382), SP24 (AY442387), SP3 (AY442366), SP13 (AY442376), SP9 (AY442372), span129 (HTU10265), SP11 (AY442374).

The GenBank accession numbers of the new HTLV-2 LTR sequences included in the phylogenetic analysis are as follows: HT567 (KF146724), HT1223 (KF146716), HT1323 (KF146717), HT3080 (KF146718), HT3989 (KF146719), HT4553 (KF146720), HT5275 (KF146721), HT5317 (KF146722), HT5588 (KF146723), HT5747 (KF146725), HT5751 (KF146726), HT6212 (KF146727), HT6217 (KF146728), HT6234 (KF146729), HT6512 (KF146730).

#### Author Disclosure Statement

No competing financial interests exist.

#### References

- Kalyanaraman VS: A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science* 1982;218:571–573.
- Vandamme AM: Evolutionary strategies of human T-cell lymphotropic virus type II. *Gene* 2000;261:171–180.
- Araujo A: Human T-lymphotropic virus type II and neurological disease. *Ann Neurol* 2004;56:10–19.
- Roucoux DF: The epidemiology and disease outcomes of human T-lymphotropic virus type II. *AIDS Rev* 2004;6:144–154.
- Hall WW: Multiple isolates and characteristics of human T-cell leukemia virus type II. *J Virol* 1992;66:2456–2463.
- Vandamme AM: African origin of human T-lymphotropic virus type 2 (HTLV-2) supported by a potential new HTLV-2d subtype in Congolese Bambuti Efe pygmies. *J Virol* 1998;72:4327–4340.
- Eiraku N: Identification and characterization of a new and distinct molecular subtype of human T-cell lymphotropic virus type 2. *J Virol* 1996;70:1481–1492.
- Alcantara LC: Brazilian HTLV type 2a strains from intravenous drug users (IDUs) appear to have originated from two sources: Brazilian Amerindians and European/North American IDUs. *AIDS Res Hum Retroviruses* 2003;19:519–523.
- Jeanmougin F: Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998;23:403–405.
- BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999;41:95–98.
- Swofford DL: *Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. PAUP\* 2002, 1–142.
- Geneious 6.1 created by Biomatters Ltd., 2013.
- Tamura K: MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
- Merezak C: Suboptimal enhancer sequences are required for efficient bovine leukemia virus propagation in vivo: Implications for viral latency. *J Virol* 2011;75:6977–6988.

Address correspondence to:

Luiz Carlos Junior Alcantara  
LHGB/CPqGM/FIOCRUZ  
Rua Waldemar Falcão 121  
Brotas, Salvador  
Bahia 40295-001  
Brazil

E-mail: lalcan@bahia.fiocruz.br