



**FUNDAÇÃO OSWALDO CRUZ**

**INSTITUTO GONÇALO MONIZ**

**Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina  
Investigativa**

**TESE DE DOUTORADO**

**CARACTERIZAÇÃO GENÔMICA PROVIRAL, CLÍNICA E EPIDEMIOLÓGICA  
DA INFECÇÃO PELO VÍRUS LINFOTRÓPICO DE CÉLULAS T HUMANAS DO  
TIPO 1 (HTLV-1)**

**THESSIKA HIALLA ALMEIDA ARAÚJO**

**Salvador – Brasil**

**2016**

**FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO GONÇALO MONIZ**

**Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina  
Investigativa**

**CARACTERIZAÇÃO GENÔMICA PROVIRAL, CLÍNICA E EPIDEMIOLÓGICA DA  
INFECÇÃO PELO VÍRUS LINFOTRÓPICO DE CÉLULAS T HUMANAS DO TIPO 1  
(HTLV-1)**

**THESSIKA HIALLA ALMEIDA ARAÚJO**

Tese apresentada ao Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa para a obtenção do grau de Doutor.

Orientador: Prof. Dr. Luiz Carlos Júnior  
Alcântara

Coorientador: Profa. Dra. Aline Miranda  
Mascarenhas

**Salvador – Brasil**

**2016**

" CARACTERIZAÇÃO GENÔMICA PROVIRAL, CLÍNICA E EPIDEMIOLÓGICA DA INFECÇÃO PELO  
VÍRUS LINFOTRÓPICO DE CÉLULAS T HUMANAS DO TIPO I (HTLV-1)."

**THESSIKA HIALLA ALMEIDA ARAÚJO**

FOLHA DE APROVAÇÃO

COMISSÃO EXAMINADORA



---

Dra. Dalila Luciola Zanette  
Pesquisadora  
IGM/FIOCRUZ



---

Dra. Ana Carolina Paulo Vicente  
Pesquisadora Titular  
IOC/FIOCRUZ



---

Dra. Isadora Cristina de Siqueira  
Pesquisadora  
IGM/FIOCRUZ

## AGRADECIMENTOS

A Deus, meu companheiro das horas difíceis e felizes, meu amigo, minha força, minha fé, minha eterna esperança...

Aos meus pais, Joel e Dilene, meus grandes exemplos de ética, generosidade, fé e amor ao próximo.

As minhas irmãs, por todo amor, confiança e incentivo profissional.

A todos os meus amigos que mesmo não compreendendo nada de HTLV-1, estão sempre ao meu lado, sendo o meu refúgio, motivos de alegria e incentivo, em especial, a minha eterna amiga, Giovanna Pinho.

A todos da minha família pelo apoio para este passo tão importante da minha vida profissional, em especial, Nany, Patrícia e Tia Goreti.

Ao meu orientador Dr. Luiz Carlos Junior Alcântara por todo conhecimento compartilhado, por acreditar no meu trabalho e por sempre incentivar o meu crescimento profissional. Minha gratidão!

A minha co-orientadora e grande amiga, Aline Miranda. Obrigada por todo incentivo, por todo carinho, por me ajudar a enxergar a luz, quando eu só via escuridão.

Ao Dr. Bernardo Galvão pelo apoio constante, por todos valiosos ensinamentos (científicos e da vida) e por ter sido um grande motivador na realização deste trabalho.

Aos alunos queridos de iniciação científica que fizeram parte deste trabalho, Felipe Ivanoff e Aline Dórea. Obrigada por toda ajuda, por deixar os meus dias muito mais leves e risonhos! Não esquecerei a trilha sonora de cada experimento, prometo!

Aos colaboradores deste projeto Dr. Bernardo Galvão, Dra. Lourdes Farré, Dra Achilea Bittencourt e Dr. Márcio Nunes, obrigada por toda ajuda e participação no desenvolvimento deste projeto.

À equipe do Instituto Evandro Chagas (IEC), em especial aos meus queridos amigos, Rodrigo, Clayton, Poliana, Lay e Janaína, por todo apoio no sequenciamento das amostras.

À equipe do Centro de HTLV da Escola Bahiana de Medicina e Saúde Pública, em especial aos meus grandes amigos Cláudio e Noilson por todo carinho, conselhos e ajuda em todos os momentos que eu necessitei. Vocês são muito especiais na minha vida!

A todos os meus colegas e amigos que eu tive uma grande sorte de conhecer e conviver no Centro de Pesquisa Gonçalo Moniz nestes 10 anos de FIOCRUZ. Em especial aos meus grandes incentivadores e amigos, Fernanda (Códon), Tika (Bina), Filipe (Lipinho), Lua, Jell, Iuka e Gisa. Obrigada por todo apoio, por ficarem ao meu lado nos momentos mais difíceis da minha vida pessoal e científica. Serei sempre grata por todas as palavras sinceras, pelos ensinamentos e por todo amor que reina neste grupo de pessoas tão generosas e especiais. Agradeço sempre a Deus por tê-los por perto.

A todos do LHGB, em especial aos “meninos” da informática, Jaq, Rukete, Fê e a nossa italiana mais brasileira de todas, Martita!

A todos os professores do Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa que participaram da minha formação acadêmica e profissional.

Aos funcionários do CPqGM pelo suporte no desenvolvimento deste projeto, em especial a Lumara, por toda gentileza e presteza usual.

A equipe da biblioteca do CPqGM pelo suporte no desenvolvimento e na formatação desta tese.

A todos que contribuíram direta ou indiretamente para a realização deste trabalho e para a minha formação profissional.

*“Porque a vida é fugaz,  
tão veloz, tão passageira.  
A gente sofre demais,  
por bobagens, por besteira.  
Tudo um dia se desfaz  
mesmo que queira ou não  
queira.  
Importa é viver em paz  
pois quando olhamos pra trás  
lá se foi a vida inteira”*

*Jenário de Fátima*

ARAÚJO, Thessika Hialla Almeida. Caracterização genômica proviral, clínica e epidemiológica da infecção pelo linfotrófico de células T do tipo 1 (HTLV-1). 106 f. il. Tese (Doutorado) – Fundação Oswaldo Cruz, Instituto Gonçalo Moniz, Salvador, 2016.

## RESUMO

O vírus linfotrófico de células T humanas tipo 1 foi o primeiro retrovírus humano isolado em 1980. Este agente etiológico está associado principalmente à paraparesia espástica tropical/mielopatia associada ao HTLV (TSP/HAM), à Leucemia/Linfoma de células T do Adulto (ATL) e à dermatite infecciosa associada ao HTLV-1 (DIH). Estima-se que 5 a 10 milhões de indivíduos estejam infectados pelo HTLV-1 em todo o mundo. Apesar da importância clínica e epidemiológica do HTLV-1, há um número limitado de genomas completos disponíveis, cerca de 0,12 genomas completos por 10.000 indivíduos infectados, somado à falta de estudos relacionados às sequências (totais e parciais) disponíveis nos bancos de dados públicos. Deste modo, este estudo foi dividido em três etapas: a primeira refere-se à avaliação do perfil e das informações disponíveis nas sequências publicadas mundialmente. Entre as sequências publicadas, 93% apresentaram informação sobre origem geográfica e apenas 39% informação sobre os perfis clínicos. Ficou clara a escassez destas informações, que são importantes para trabalhos envolvendo a diversidade viral, estudos clínicos e epidemiológicos. Estes resultados direcionaram a realização das etapas seguintes: (i) o desenvolvimento de um protocolo no formato de capítulo de livro, que ensina de forma didática os pesquisadores interessados nos estudos relacionados à epidemiologia molecular e genotipagem do HTLV-1, a ferramenta *HTLV-1 Subtyping Tool* e o *HTLV-1 Molecular Epidemiology Database* foram abordados neste capítulo. (ii) Na última etapa, foram gerados 31 genomas completos provenientes de indivíduos com diferentes perfis clínicos através da plataforma de sequenciamento *Ion Torrent*. Essas sequências foram montadas e analisadas, sendo todas pertencentes ao subtipo Cosmopolita, subgrupo transcontinental e na análise da distância genética, as sequências apresentaram uma baixa diversidade intra e intergrupos. Posteriormente, utilizamos o programa *Geneious R6* para avaliar a região promotora viral quanto à identificação de sítios de ligação de fatores de transcrição. Identificamos duas mutações que levaram a perda do sítio de ligação do fator Sp1. O mesmo programa foi utilizado para identificar variantes nas regiões codificantes do genoma. Não foram encontradas associações entre as variantes e os diferentes perfis clínicos dos pacientes (TSP/HAM, DIH, ATL, Assintomáticos). A análise físico-química demonstrou que uma mutação nos genes *env* e *gag*, resultou em sequências proteicas (gp46 e p19) com menor antigenicidade. Em contrapartida, uma mutação no gene *gag* resultou em uma maior antigenicidade em p24. Em relação a análise das modificações pós-traducionais, identificou-se que pequeno número de mutações relacionadas a criação e/ou perda de sítios nos diferentes perfis clínicos, sem associação entre eles. Em uma segunda fase do estudo, buscamos as mutações encontradas nas sequências geradas, em sequências previamente publicadas no Genbank, corroborando com os dados anteriores, os resultados não mostraram significância estatística. Este estudo contribuiu para o aumento do número de genomas completos disponíveis. Além disso, para investigar melhor a contribuição das mutações do HTLV-1 para o desfecho da doença é necessário avaliar a interação do genoma viral e as características do hospedeiro humano

**Palavras-chaves:** Vírus Linfotrófico de Células T Humanas tipo 1, Epidemiologia molecular do HTLV-1, Genoma completo

ARAÚJO, Thessika Hialla Almeida. Epidemiological, clinical and genomic features of human T-cell lymphotropic virus type-1 (HTLV-1) infection. 106 f. il. Tese (Doutorado) – Fundação Oswaldo Cruz, Instituto Gonçalo Moniz, Salvador, 2016.

## ABSTRACT

The Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) is the first human retrovirus isolated in 1980 and is associated with several etiologies as TSP/HAM (Tropical Spastic Paraparesis/HTLV-1-Associated Myelopathy), ATL (Adult T-cell leukemia/lymphoma) and HID (HTLV-1 associated infective dermatitis). It is estimated that approximately 5 to 10 million people are infected with HTLV-1 and this infection is worldwide distributed. Despite the clinical and epidemiological importance of HTLV-1 infection, there is a limitation of number of complete genomes available, about 0.12 complete genomes per 10,000 infected individuals and few studies related to the sequences (total and partial) available in the public databases. This study was performed in three stages: the first one refers to the analysis of pattern and characterization about information associated to sequences published worldwide. Among all HTLV-1 available sequences, 93% of them have information about geographic origin, and only 39% are associated to information about the clinical status of host infected individuals: 29% of them were originated from individuals with TSP/HAM and 67,8% were generated from asymptomatic individuals. The relative scarcity of data of available sequences, may affect the studies involving viral diversity, clinical and epidemiological investigations. These results led to the following steps: (i) the development of a protocol in the book chapter format, which teaches didactically the researchers interested in studies related to the molecular epidemiology and genotyping of HTLV-1, the HTLV-1 tool Subtyping Tool and the HTLV-1 Molecular Epidemiology Database have been covered in this chapter. (ii) At the third stage, were generated and characterized 31 HTLV-1 complete genomes sequences derived from individuals with different clinical status through the Ion Torrent sequencing platform. These sequences were assembled and analyzed. All the sequences were genotyped as Cosmopolitan subtype, Transcontinental subgroup. The genetic distance, showed low intragroup and intergroup diversity. The Geneious R6 software was used to analyze the non-coding region (LTR) to identify the transcriptions factors binding sites. Our analyses demonstrated that two LTR mutations were able to abolish the binding of Sp1 (transcription factor). The same software was used to identify the variants at genome coding regions. No statistical relationship between variants and different clinical profiles (TSP/HAM, HID, ATL, Asymptomatic) were detected. The physico-chemical analysis showed that one mutation at env and gag genes was able to decrease the antigenicity while other mutation at gag gene (p24) was able to increase the antigenicity. The post-translational modification analysis demonstrated a low frequency of mutations associated with the creation or abrogation of these sites in different clinical profiles, with no statistical association. In a second phase of the study, we search the mutations found in the sequences generated in sequences previously published in Genbank, corroborating previous data, there was no statistical significance. This study contributed to increase of HTLV-1 complete genomes in world. Furthermore, to investigate better the contribution of HTLV-1 mutations for the disease outcome it is necessary evaluate the interaction of the viral genome and characteristics of the human host.

**Key Words:** Human T-Cell Lymphotropic Virus Type 1, HTLV-1 molecular epidemiology, complete genome.



## LISTA DE FIGURAS

<b>Figura 1.</b> Estrutura do HTLV-1.....	18
<b>Figura 2.</b> Organização Genômica do HTLV-1.....	19
<b>Figura 3.</b> Principais vias de dispersão do HTLV-1.....	24
<b>Figura 4.</b> Formação de Conduítes .....	24
<b>Figura 5.</b> Distribuição do HTLV-1 no mundo.....	27
<b>Figura 6.</b> Prevalência de HTLV-1/2 entre doadores de sangue em capitais de 26 estados brasileiros e no Distrito Federal .....	28
<b>Figura 7.</b> Representação de um indivíduo com TSP/HAM. ....	33
<b>Figura 8.</b> Indivíduo com a forma clínica indolente da ATL apresentando lesões na pele. ....	37
<b>Figura 9.</b> Dermatite infecciosa em uma criança infectada pelo HTLV-I. Lesões eczematosas observadas predominantemente no couro cabeludo, face e região auricular.....	40
<b>Figura 10.</b> Análise comparativa entre a média de indivíduos infectados pelo HTLV-1 nas regiões endêmicas versus o número de genomas completos disponíveis. ....	43

## LISTA DE TABELAS e QUADROS

<b>Tabela 1.</b> Proteínas do HTLV-1 e suas funções.....	21
<b>Quadro 1.</b> Critérios de Diagnóstico de TSP/HAM.....	32
<b>Quadro 1.</b> Formas clínicas da ATL e suas principais características específicas.....	36

## LISTA DE ABREVIATURAS E SIGLAS

AMPC: Monofosfato cíclico de adenosina (*Cyclic adenosine monophosphate*)

APOBEC3G - *apolipoprotein B-editing catalitic polypeptide 3G*

AP-1 – Proteína ativadora 1 (*Activator protein 1*)

ATL – Leucemia/linfoma de células T do adulto (*Adult T-cell leukemia-lymphoma*)

CA – Capsídeo viral

CD4+ – Linfócitos T CD4+

CD8+ – Linfócitos T CD8+

CPV – Carga proviral

CREB/ATF – Proteína de ligação ao elemento de resposta ao cAMP (*cAMP response element-binding protein/ Activating transcription factors*)

CTL – Linfócito T citotóxico

DC – Célula dendrítica (*dendritic cell*)

DIH – Dermatite infecciosa associada ao HTLV-1

DNA – Ácido desoxirribonucléico

E Box – *Enhancer*

ELISA – *Enzyme-Linked Immunosorbent Assay*

env – Envelope

gag – Grupo antigênico

GLUT-1 – Molécula transportadora de glicose tipo 1 (*Glucose transporter 1*)

gp21 – Glicoproteína transmembrana

gp46 – Glicoproteína de superfície

HBZ – *HTLV-1 bZIP factor gene*

HIF-1 – Fator de indução de hipóxia tipo 1

HIV-1 – Vírus da imunodeficiência humana do tipo 1

HLA – Antígeno leucocitário humano (*Human leukocyte antigen*)

HSPG – Heparan Sulfato-Proteoglicano

HTLV-1 – Vírus linfotrópico de células T humanas tipo 1 (*human T-cell lymphotropic virus type 1*)

HTLV-2 – Vírus linfotrópico de células T humanas tipo 2 (*human T-cell lymphotropic virus type 2*)

HTLV-3 – Vírus linfotrópico de células T humanas tipo 3 (*human T-cell lymphotropic virus type 3*)

HTLV-4 – Vírus linfotrópico de células T humanas tipo 4 (*human T-cell lymphotropic virus type 4*)

IN – Integrase

LEF1 – *lymphoid-enhancer binding fator 1*

LCR – Líquido cefalorraquiano

LTR – *Long terminal repeats*

NC – Nucleocapsídeo

NF- kB – Nuclear fator kB

Nm – nanômetros

NRP-1 – Neurofilina-1

ORF – Fase de leitura aberta (*Open Reading Frame*)

p12, p30 e p13 – Proteínas acessórias (proteínas de 12, 30 e 13 kilodaltons, respectivamente).

p15 – Proteína do nucleocapsídeo de 15 kD (kilodaltons)

p19 – Proteína da matriz de 19 kD (kilodaltons)

p24 – Proteína do capsídeo de 24 kD (kilodaltons)

RNA – Ácido ribonucleico

RNA<sub>m</sub> – Ácido ribonucleico mensageiro

SAMHD1 - *SAM domain- and HD domain-containing protein 1*

SNC – Sistema Nervoso Central

SRF – Fator de resposta ao soro (*Serum Response Factor*)

SU – Proteína de superfície

TCF1 - *T-cell factor 1*

TM – Proteína transmembrana

TR – Transcriptase reversa

tRNA – Ácido ribonucleico transportador

TSP/HAM – Paraparesia Espástica Tropical/ Mielopatia Associada ao HTLV-1  
(*Tropical Spastic Paraparesis/ HTLV-I-Associated Myelopathy*)

WB – *Western Blot*

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	16
1.1	Vírus Linfotrópico de Células T Humanas (HTLV).....	16
1.2	Morfologia do HTLV .....	17
1.3	Estrutura viral do HTLV-1 e seus produtos proteicos.....	18
1.4	Ciclo de replicação viral e mecanismos de infecção .....	21
1.5	Variabilidade Genética Viral .....	25
1.6	Distribuição geográfica da infecção pelo HTLV-1 .....	26
1.7	Epidemiologia molecular do HTLV-1.....	28
1.8	Diagnóstico sorológico e molecular .....	29
1.9	Vias de transmissão do HTLV-1 .....	31
<b>2</b>	<b>DOENÇAS ASSOCIADAS AO HTLV-1</b> .....	32
2.1	Paraparesia Espástica Tropical/Mielopatia Associada ao HTLV-1 (TSP/HAM).....	32
2.2	Leucemia/Linfoma de Células T do Adulto (ATL).....	35
2.3	Dermatite Infecciosa associada ao HTLV-1.....	39
<b>3</b>	<b>JUSTIFICATIVA</b> .....	40
<b>4</b>	<b>OBJETIVOS</b> .....	45
4.1	Geral.....	45
4.2	Específicos.....	45
<b>5</b>	<b>RESULTADOS</b> .....	46
5.1	Artigo I: Inferências sobre o cenário global da infecção pelo vírus linfotrópico de células T humanas do tipo 1, utilizando data mining de sequências virais. ....	47
5.2	Capítulo de livro: Banco de dados, <i>data mining</i> e genotipagem do HTLV-1 .....	57
5.3	Artigo 2 - <i>manuscrito em preparação</i> : Caracterização molecular de novos genomas completos de HTLV-1 provenientes de pacientes com diferentes perfis clínicos.....	67
<b>6</b>	<b>DISCUSSÃO</b> .....	88
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	93
	<b>REFERÊNCIAS</b> .....	94

## 1 INTRODUÇÃO

### 1.1 Vírus Linfotrópico de Células T Humanas (HTLV)

Há quase quatro décadas, Poiesz e colaboradores (1980) isolaram e descreveram o primeiro retrovírus humano a partir da investigação de um paciente com linfoma cutâneo de células T. Entretanto, anos antes no sudoeste do Japão, pacientes já tinham sido diagnosticados com uma forma distinta de leucemia que possuía características clínicas e morfologia celular especiais, nomeada de Leucemia/Linfoma de células T do Adulto (ATL) (UCHIYAMA et al., 1977). Os soros destes pacientes foram posteriormente analisados, encontrando-se anticorpos anti-HTLV-1. Estes resultados foram fundamentais para associar o então vírus descrito HTLV-1 (Vírus linfotrópico da célula T humana do tipo 1) às células T malignas da ATL (GALLO, 1981).

Estudos subsequentes identificaram e descreveram mais três tipos virais, o HTLV-2 (KALYANARAMAN et al., 1982), HTLV-3 (CALATTINI et al., 2005; WOLFE et al., 2005) e HTLV-4 (WOLFE et al., 2005). Compartilhando 65% de similaridade na sequência de nucleotídeos com o HTLV-1 (SODROSKI, 1992), o HTLV-2 foi isolado de um paciente com uma forma atípica de Leucemia de células T pilosas, entretanto, não são descritos na literatura casos de leucemia associada a este tipo específico do HTLV, conquanto uma linfocitose benigna e raros eventos de alterações neurológicas são descritas nos indivíduos infectados (HJELLE et al., 1992; DOONEIEF et al., 1996; FEUER; GREEN, 2005; BISWAS et al., 2009).

O HTLV-3 e HTLV-4 foram isolados de indivíduos provenientes de Camarões e da África Central (WOLFE et al., 2005; CALATTINI et al., 2005), estes dois tipos virais não possuem informações suficientes para associá-los a doenças, em indivíduos infectados (WOLFE et al., 2005).

O HTLV-1 é o principal tipo de interesse biomédico, desde à sua primeira associação no desenvolvimento da leucemia/linfoma de células T de adulto no início da década de 80 (POIESZ et al., 1980), aos casos descritos de paraparesia espástica tropical (*Tropical Spastic Paraparesis-TSP*) e mielopatia associada ao HTLV-1 (*HTLV-1 Associated Myelopathy-HAM*), na Martinica e no Japão, respectivamente. Estudos mostraram se tratar da mesma etiologia, sendo o HTLV-1 o agente diretamente

relacionado a esta síndrome neurológica denominada TSP/HAM (ROMÁN; OSAME, 1988).

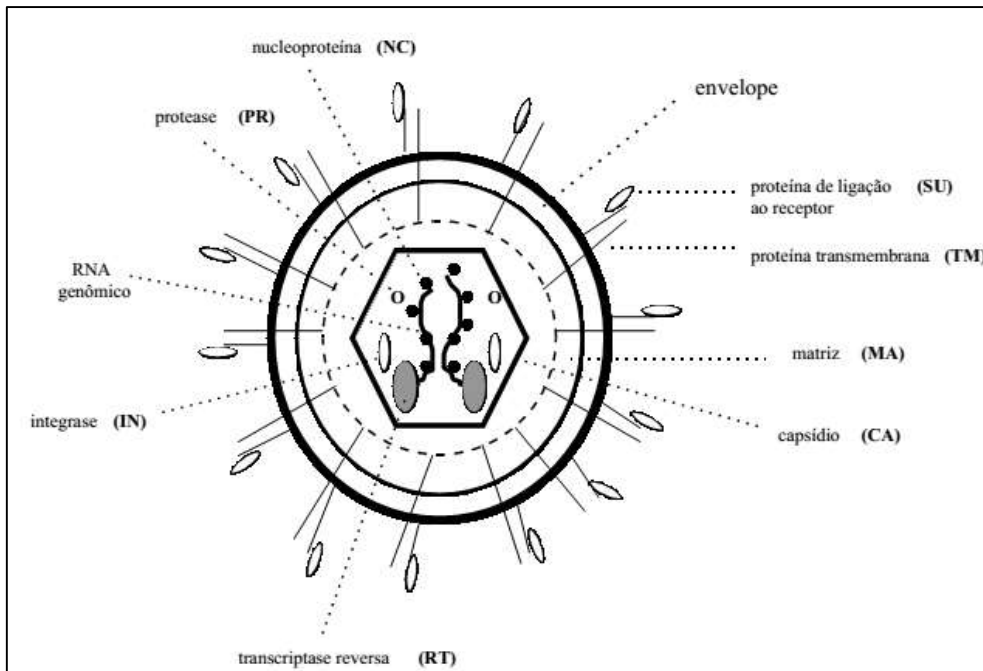
Posteriormente, vários estudos associaram o HTLV-1 com outras importantes patologias, tais como polimiosites (MORGAN et al., 1989), artropatias (NISHIOKA, 1996), uveítes (MOCHIZUKI et al., 1992), estrogilodíase (NAKADA et al., 1984) e dermatites infecciosas (LA GRENADE, 1996).

## 1.2 Morfologia do HTLV

O HTLV pertence à família Retroviridae, à subfamília Orthoretrovirinae, gênero Deltaretrovirus. Similar a outros retrovírus, o HTLV tem uma morfologia esférica a pleomórfica, formado basicamente por um envelope (diâmetro entre 100 a 140 nanômetros), uma matriz proteica e um nucleocapsídeo icosaédrico (cerca de 80 a 100 nm).

O envelope é composto por um complexo proteico glicosilado: uma proteína de superfície (SU) extracelular e uma proteína transmembrana (TM), que atravessa esta estrutura e ancora a SU. Logo após a membrana do envelope encontra-se a matriz viral que é composta pela proteína da matriz de 19 kD (kilodaltons), ou p19. O capsídeo (CA) constitui o núcleo da partícula viral e é composto principalmente por uma proteína de 24 kD (p24). No interior deste núcleo, localiza-se o genoma viral composto por duas fitas simples de RNA (ácido ribonucléico) com polaridade positiva (BURKE, 1997), às quais estão associadas a várias proteínas básicas, chamadas de proteínas do nucleocapsídeo (NC) de 15 kD (p15). Outras proteínas também estão presentes no interior do CA, como a transcriptase reversa (TR) e a integrase (IN), essenciais no processo de integração do DNA no genoma da célula hospedeira. Além disso, a partícula viral possui proteínas envolvidas em eventos catalíticos, tais como protease e RNase H (Figura 1).



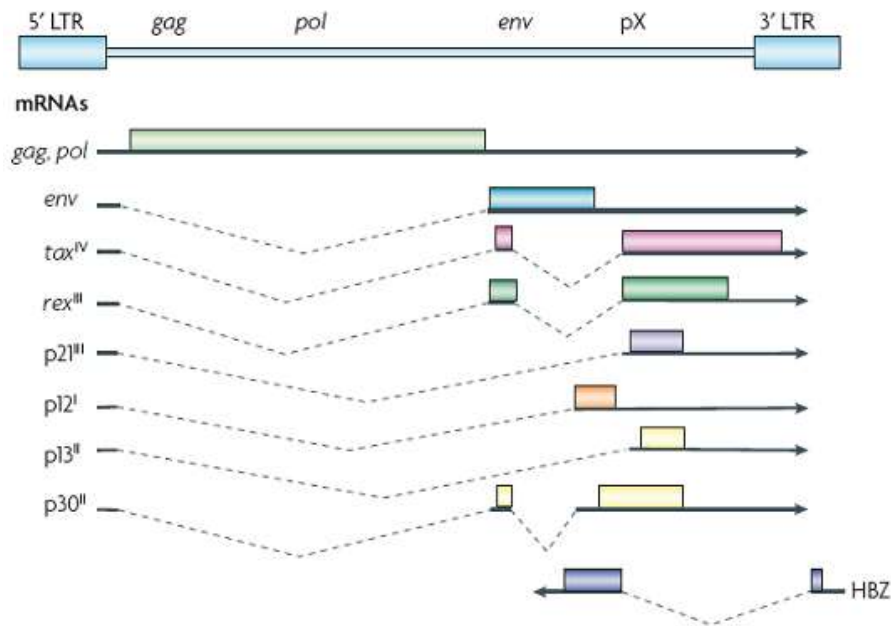


**Figura 1.** Estrutura do HTLV-1. (Adaptado do *HTLV – Cadernos Hemominas*, 2015).

### 1.3 Estrutura viral do HTLV-1 e seus produtos proteicos

O HTLV-1, HTLV-2, HTLV-3 e HTLV-4 possuem o DNA proviral de 9032, 8952, 8553 e 8791 pares de bases (pb), respectivamente (SEIKI et al., 1983; TSUJIMOTO et al., 1988; CALATTINI et al., 2005; WOLFE et al., 2005).

O genoma do HTLV-1 é composto por duas fitas simples de RNA é formado pelos genes *gag* (grupo antigênico), *pol* (polimerase), *env* (envelope) e região *pX*, além de ser flanqueado por duas regiões repetidas, chamadas *LTR* (*long terminal repeats*), responsável pela integração do DNA proviral no DNA cromossômico do hospedeiro e pela regulação transcricional do genoma do HTLV-1 (GREEM; CHEN, 2001) (Figura 2).



**Figura 2.** Organização Genômica do HTLV-1. Desenho esquemático adaptado de MATSUOKA e JEANG, 2007.

As proteínas estruturais do capsídeo são codificadas pelo gene *gag*. Uma poliproteína precursora é formada e posteriormente clivada dando origem às proteínas da matriz (p19), do capsídeo (p24) e nucleocapsídeo (p15).

As enzimas virais transcriptase reversa (TR), RNase H e integrase (IN) são codificadas pelo gene *pol* e são responsáveis respectivamente pela síntese do DNA viral a partir do seu genoma RNA, estando presente no cerne da partícula viral; remoção da fita RNA molde após a síntese da cadeia de DNA, degradando seletivamente o RNA da molécula híbrida DNA-RNA e integração do DNA viral no genoma da célula hospedeira.

O gene *env* codifica as proteínas do envelope viral (Env). A proteína precursora do Env é clivada para gerar os produtos maduros, a glicoproteína de superfície (gp46 – SU) e uma proteína transmembrana (gp21- TM) (SHIMOTOHNO et al., 1985) ambas são importantes para a interação com a célula alvo e posterior infecção.

A enzima protease é codificada pelo desvio de leitura dos genes *gag* e *pol*, atuando sobre as cadeias poliproteicas, clivando-as para a formação das proteínas estruturais maduras encontradas na partícula viral.

A região situada imediatamente antes da região *LTR* 3' (*long terminal repeat*), denominada *pX*, contém 4 *Orfs* (*Open Reading Frames*) que codificam proteínas regulatórias (Tax, Rex, HBZ) e acessórias (p12/p8, p13, p30). A proteína Tax atua na

modulação da expressão viral, recrutando fatores de transcrição para a região 5' *LTR* e em importantes vias de processos de proliferação, ciclo celular e apoptose, sendo responsável pela ativação dos fatores de transcrição pertencentes à família *CREB/ATF* (*cAMP response element-binding protein/ Activating transcription factors*) e à família *AP-1* (*activator protein 1*); ativação do fator de resposta ao soro (SRF), além de interagir com *NF-κB* (*nuclear fator κB*), atuando diretamente como reguladores de ciclo celular e *checkpoints*. Enquanto que a proteína Rex, regula a expressão gênica viral em nível postranscricional, a proteína HBZ (*HTLV-1 bZIP factor*) que é codificada por um RNA mensageiro anti-sense, sequestra fatores de transcrição, o que determina à diminuição da transcrição do genoma viral mediada por Tax. Além disso, a HBZ reprime a via clássica de sinalização de NF-κB, estando relacionada nas etapas finais da oncogênese relacionada à infecção pelo HTLV-1 (ZHI et al., 2011).

As duas regiões LTR idênticas, localizadas nas extremidades 5' e 3' do DNA proviral são formadas por três subunidades: U3, R e U5. A região U3 exerce um papel importante na infecção viral, por conter três elementos importantes, tais como o domínio TATA box (TATAA), que antecede o sítio de iniciação da transcrição; três sequências repetitivas imperfeitas, conhecidas como os elementos responsivos à proteína Tax (*Tax responsive elemento* - TxRE/TRE), importantes para transativação mediada por Tax (BRADY et al., 1987) e o sinal de poliadenilação (AATAAA), fundamental no processo de transcrição (MCDEVITT et al., 1986).

A transcrição gera três moléculas de mRNA: o mRNA genômico, utilizado para a síntese dos produtos dos genes *gag* e *pol*, transcrito da extremidade *LTR* 5' até a junção R-U5 na extremidade *LTR* 3'; o mRNA sub-genômico, sintetizado a partir de uma única etapa de processamento e codificante do produto do gene *env*; e um segundo mRNA subgenômico, duplamente processado através da remoção de dois íntrons, que codifica as proteínas regulatórias Tax e Rex com, pelo menos, quatro fases de leitura aberta (*Open Reading Frame- ORF*) (Tabela 1).

Na análise do genoma total, o gene *env* e a região *LTR* possuem uma maior variabilidade genética (SHERMAN et al., 1992; SWITZER et al., 2009), este atributo é fundamental para os estudos de epidemiologia molecular do HTLV, elucidando as teorias sobre a origem e disseminação do vírus e dos processos migratórios de populações humanas infectadas.

**Tabela 1.** Proteínas do HTLV-1 e suas funções.

<b>HTLV-1 proteínas e glicoproteínas</b>	<b>Funções</b>
<b>Proteínas do envelope (codificadas por env)</b>	
Proteína de superfície (gp46)	Liga-se ao receptor da célula hospedeira
Glicoproteína transmembrana (gp21)	Ancora glicoproteínas de superfície ao vírus
<b>Proteínas estruturais (codificadas por gag)</b>	
Proteína da matriz (p19)	Organiza componentes virais na membrana celular interna.
Proteína do Capsídeo (p24)	Protege RNA viral e proteínas
Proteína do Nucleocapsídeo (p15)	Proteína que liga os ácidos nucleicos
<b>Proteínas Funcionais (codificadas por pol)</b>	
Protease (p14)	Cliva poliproteínas a componentes funcionais
Transcriptase reversa (p95)	Converte RNA de fita única em DNA de fita dupla
Integrase	Facilita a inserção do provírus no DNA da célula hospedeira
<b>Proteínas reguladoras (codificadas por pX)</b>	
Tax (p40)	<ul style="list-style-type: none"> <li>• Ativa a transcrição do provírus.</li> <li>• Ativa a transcrição de uma ampla variedade de genes do hospedeiro.</li> <li>• Desregula fatores do ciclo celular e imortaliza as células T.</li> </ul>
Rex	Modula o transporte de RNA viral
p12 <sup>I</sup>	Replicação viral e ativação de células T
p27 <sup>I</sup>	Desconhecida
p13 <sup>II</sup>	Interfere na proliferação e transformação celular, promove apoptose e seu alvo é a mitocôndria.
p30 <sup>II</sup>	<ul style="list-style-type: none"> <li>• Promove a transcrição viral e proliferação celular.</li> <li>• Desregula múltiplos genes envolvidos nos efeitos pré-apoptóticos e antiapoptóticos.</li> </ul>
<i>HBZ</i>	Suprime regulação da transcrição viral e auxilia na proliferação de células T

(Adaptado do HTLV -Cadernos Hemominas, 2015).

#### 1.4 Ciclo de replicação viral e mecanismos de infecção

No processo de reconhecimento e conseqüente infecção pelo HTLV-1, as células T CD4<sup>+</sup> são os principais alvos (RICHARDSON et al., 1990). Entretanto, estudos demonstraram a presença do HTLV-1 em outros tipos celulares, tais como monócitos e células B (KOYANAGI et al., 1993), células dendríticas (KNIGHT et al., 1993), endoteliais (SETOYAMA et al., 1998), células T CD8<sup>+</sup> (HANON et al., 2000) e macrófagos (NATH et al., 2002).

A partir da interação específica entre as glicoproteínas do envelope viral (gp46) e receptores celulares, ocorre a entrada do vírus na célula hospedeira. Moléculas tais como a neurofilina-1 (NRP-1), que constitui uma glicoproteína de membrana (GHEZ et al., 2006) e os proteoglicanos de heparan sulfato (HSPGs) são usadas na ligação inicial com a célula (PINON et al., 2003), além do transportador de glicose-1 (GLUT-1) que também participa do processo de entrada do vírus no citoplasma e na fusão na membrana plasmática (BATTINI et al., 2003; MANEL et al., 2004; COSKUN; SUTTON, 2005). Basicamente, após a ligação do peptídeo de fusão, que compreende os sítios 300 a 331 da gp46 às moléculas de HSPG na superfície celular, aumenta-se a adesão de NRP-1, assim, forma-se um complexo que propicia a exposição dos domínios de ligação em GLUT-1, no qual a gp46 finalmente se liga, iniciando o processo de fusão das membranas celular e viral, permitindo desta forma, a entrada do genoma viral na célula hospedeira. Este processo, também depende de modificações conformacionais da glicoproteína transmembrana (gp21) (LAMB et al., 2008; JONES et al., 2011).

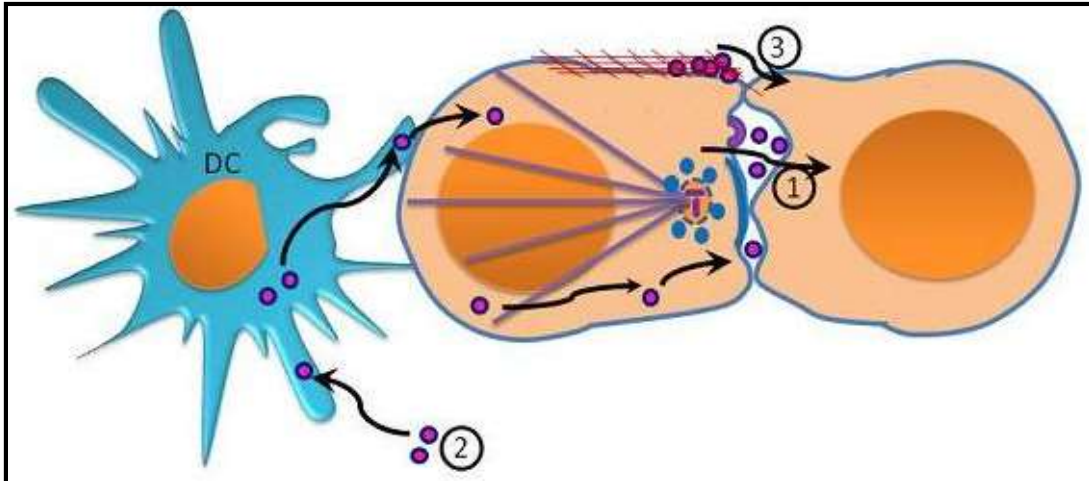
Após a fusão, ocorre o desnudamento do capsídeo e todo o material viral é liberado no citoplasma. Posteriormente, sucede a transcrição do RNA viral em DNA pela RT (transcriptase reversa) utilizando como iniciador um tRNA (RNA transportador) específico. Em seguida, o DNA que é transportado para o núcleo, integra-se ao genoma do hospedeiro pela ação da integrase viral, sendo considerado a partir deste momento, um provírus. Este processo marca a finalização da fase precoce do ciclo de multiplicação viral e o início da fase tardia, caracterizada pela síntese do RNA viral, mediada por enzimas do hospedeiro (SEIKI et al., 1984). A síntese do RNA viral leva à formação de um longo transcrito primário, o qual é processado para formar as moléculas de RNA mensageiro (mRNA) e o RNA genômico. Para a montagem da partícula viral, as proteínas são sintetizadas nos ribossomos e algumas sofrem modificações pós-traducionais.

Os precursores Gag e Gag-Pro-Pol são processados e relacionados à montagem e brotamento, não sendo clivados até a montagem da partícula viral. Para

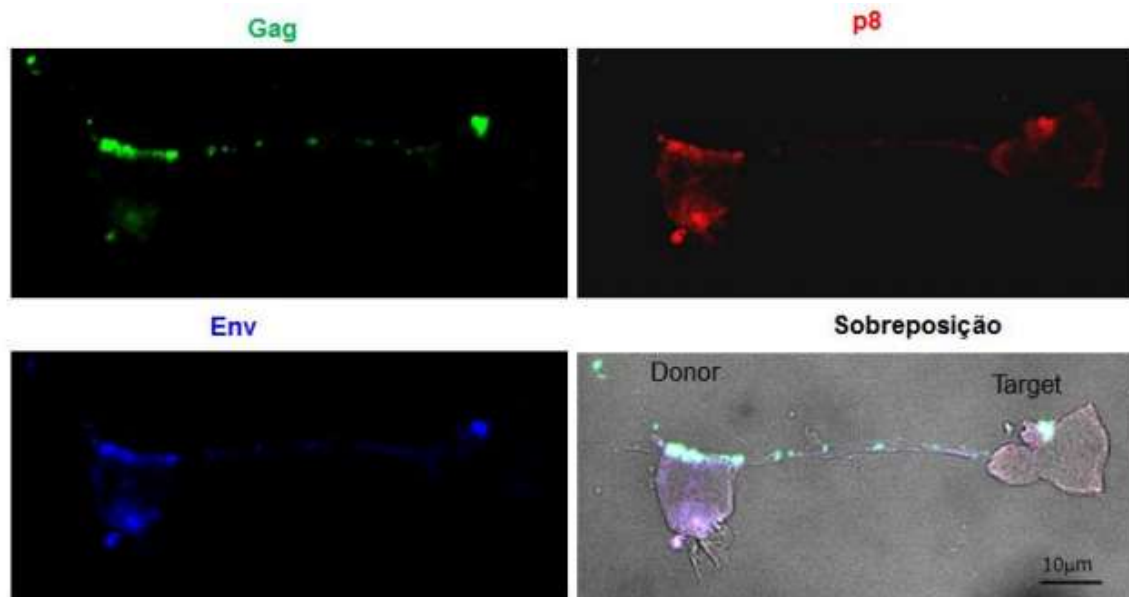
a formação da partícula infecciosa faz-se necessário o processamento proteolítico da proteína do CA (p24) e finalmente, forma-se a partícula viral madura. Estas novas partículas virais, em modelos clássicos de replicação dos retrovírus, brotam da superfície celular para reiniciar o ciclo em outra célula hospedeira.

Diferente do HIV (*Human Immunodeficiency Virus*) por exemplo, estudos demonstram outros eventos relacionados à replicação do HTLV-1, sendo o principal mecanismo *in vivo*, a expansão clonal das células infectadas, via mitose (WATTEL et al., 1995; CIMARELLI et al., 1996). Ademais, o HTLV-1 é capaz de induzir eventos de polarização das células, permitindo a passagem de proteínas e do genoma viral, este fenômeno é conhecido como sinapse viral (BANGHAM, 2003). Assim, ocorre a formação de um centro de organização microtubular (MTOC-*Microtubule-organizing center*) que é polarizado na junção de uma célula infectada com uma outra não infectada. Nesta interconexão, há um acúmulo de proteínas codificadas a partir do gene *gag* e de material genômico (RNA), que será transportado de uma célula para a outra (célula infectada para a não infectada) (MATSUOKA; JEANG, 2007; MAJOROVITS et al., 2008) (Figura 3).

Estudos recentes identificaram modelos virais extracelulares de biofilmes, formadas basicamente de moléculas de carboidratos da matriz extracelular, proteínas de adesão a exemplo das lecitinas e altas concentrações de vírions (Figura 3) (PAIS-CORREIA et al., 2009; NEJMEDDINE; BANGHAM, 2010). Tais estruturas facilitariam a dispersão de partículas virais para uma célula não infectada, pela alta concentração destas na superfície de uma célula infectada. E em 2010, estudos conduzidos por Prooyen e colaboradores detectaram proteínas virais, como p8 e proteínas traduzidas a partir dos genes *env* e *gag*, em conduítes formados entre linfócitos infectados e linfócitos não-infectados, sugerindo, portanto, um novo modelo de disseminação viral utilizado pelo HTLV-1.



**Figura 3.** Principais vias de dispersão do HTLV-1. DC (Dendritic cells). 1- Indica a infecção pelo mecanismo de sinapse virológica, 2- Indica a infecção pelo mecanismo clássico, 3- Indica a infecção facilitada por mecanismos de formação de biofilmes na superfície da célula infectada. (Adaptado de NEJMEDDINE; BANGHAM, 2010).



**Figura 4.** Formação de Conduítes (Adaptado de VAN PROOYEN et al., 2010).

Os mecanismos de infecção virais são baseados nas ações antagônicas entre as proteínas virais e algumas proteínas do hospedeiro, que são denominadas de fatores de restrição viral. O APOBEC3G (*apolipoprotein B-editing catalytic polypeptide 3G*) é um destes fatores que se destacam na infecção do HTLV-1, agindo na supressão da replicação viral, porém, este fator é controlado pela ação da proteína viral Gag, que inibe a sua incorporação no vírion (DERSE et al., 2007).

Recentemente, estudos demonstraram que o fator de restrição viral SAMHD1 (*SAM domain- and HD domain-containing protein 1*) foi capaz de suprimir a replicação

do HTLV-1 em monócitos (SZE et al., 2013). Além dele, outros fatores têm sido relatados, tais como TCF1 (*T-cell factor 1*) e LEF1 (*lymphoid-enhancer binding factor 1*), ambos inibem a função da proteína Tax e conseqüentemente, impedem a replicação viral e proliferação de células infectadas. Além disso, esses fatores restringem o tropismo do vírus às células T periféricas, o que pode ser um processo de adaptação do HTLV-1. Se o HTLV-1 pudesse replicar de forma eficiente no timo, por exemplo, ele poderia causar sérios danos ao sistema imunológico do hospedeiro e, portanto, ao hospedeiro. Assim, uma alta taxa desses fatores é encontrada no timo, diferentemente das baixas taxas encontradas nas células T CD8+, isso é um dos fatores que possibilitam a infecção destas células, que são eficientes na transmissão sexual ou através da amamentação, possibilitando a propagação o vírus (GUANGYONG et al., 2015).

### 1.5 Variabilidade Genética Viral

O modelo de replicação por meio de expansão clonal das células infectadas, citado anteriormente, e conseqüentemente a ínfima utilização da transcriptase reversa (característico do modelo clássico da replicação) (WATTEL et al., 1995; BANGHAM; OSAME, 2005), são pontos importantes que conferem ao genoma do HTLV-1 uma excelente estabilidade genética, quando comparada a outros retrovírus (GESSAIN et al., 1996).

As taxas de substituição nucleotídicas para o gene *env* e a região *LTR* ficam em torno de  $3,44 \times 10^{-7}$  e  $6,55 \times 10^{-7}$  substituições por sítio por ano, respectivamente (LEMEY et al., 2005). Estudos demonstraram uma taxa de substituição da ordem de 1% a cada 1000 anos (SHERMAN et al., 1992; GESSAIN et al., 1996). Esta grande estabilidade do genoma do HTLV, o torna um excelente marcador molecular para traçar eventos de migração das populações humanas, bem como para o entendimento dos mecanismos da evolução viral e monitoramento dos meios de transmissão (GESSAIN et al., 1992).

A partir de uma análise do genoma proviral, as regiões *LTR* e *env* do HTLV-1 apresentam uma maior variabilidade em relação às outras regiões genômicas, tais como os genes *gag* e *pol*, que por sua vez, possuem uma alta similaridade entre diferentes isolados provirais. Esta alta variabilidade de *LTR* e do gene *env* é fundamental para as análises que objetivam a caracterização genética dos subtipos

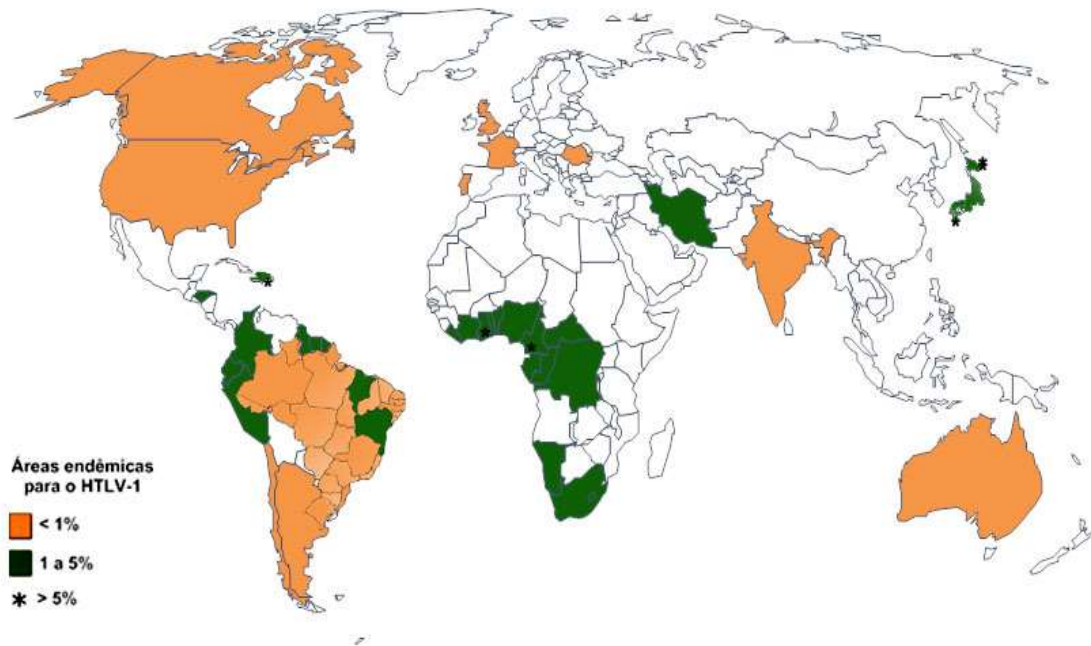


virais (SANTOS; LIMA, 2005). Considerando o HTLV-1, há divergência de 2 a 8% entre os subtipos, e a taxa de variação genética dentro do mesmo subtipo (intra-subtipo) é menor que 0,5% (DAENKE et al., 1990, GESSAIN et al., 1996).

## **1.6 Distribuição geográfica da infecção pelo HTLV-1**

Estima-se que aproximadamente cinco a dez milhões de pessoas são infectadas pelo HTLV-1 em todo o mundo (GESSAIN; CASSAR, 2012). As taxas de soroprevalência diferem, de acordo com a região geográfica, com a composição sócio-demográfica da população estudada e os comportamentos de risco individuais. Dados epidemiológicos mostram que a infecção pelo HTLV-1 tem distribuição mundial (DE THÉ; KAZANJI, 1996), no entanto, algumas áreas são endêmicas para esta infecção: sudoeste do Japão (YAMAGUCHI, 1994; MUELLER et al., 1996), África sub-Saara (GESSAIN; DE THÉ, 1996), regiões do Caribe (HANCHARD et al., 1990), áreas localizadas no Irã e Melanésia (MUELLER, 1991) e Brasil (CATALAN-SOARES et al., 2004) (Figura 5).

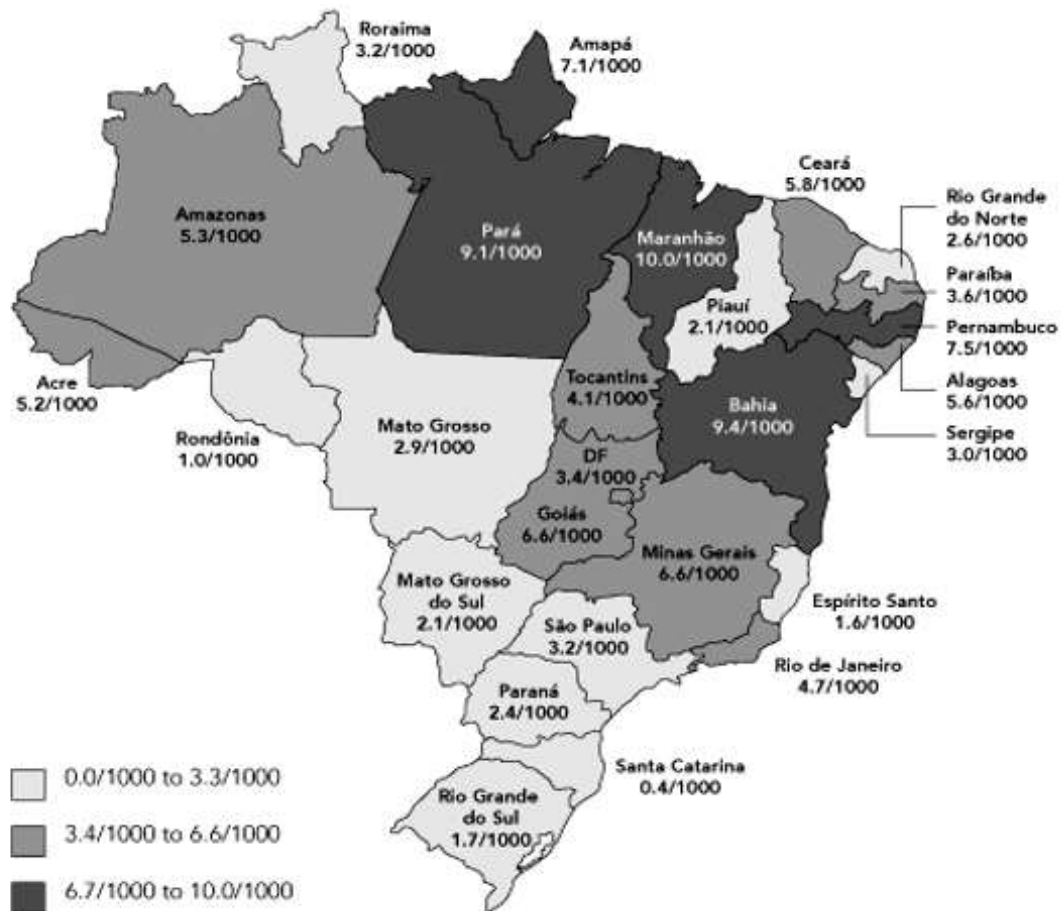
O real cenário da prevalência da infecção pelo HTLV-1 pode está subestimado, levando-se em consideração que a maioria dos estudos epidemiológicos são realizados em populações específicas, principalmente em doadores de sangue, gestantes e pacientes hospitalizados (GESSAIN; CASSAR, 2012).



**Figura 5.** Distribuição do HTLV-1 no mundo. (Adaptado de PROIETTI et al., 2005).

O Brasil é considerado uma área endêmica para o HTLV-1, porém a distribuição do vírus no território nacional é heterogênea, destacando-se os estados do Pará (Belém-9.1/1000 doadores de sangue), Maranhão (São Luís- 10.0/1000 doadores de sangue), Pernambuco (Recife-7.5/1000 doadores de sangue) e Bahia (Salvador-1,8% na população geral, ou 9.4/1000 doadores de sangue) com as maiores prevalências (DOURADO et al., 2003; CATALAN-SOARES et al., 2005) (Figura 6).

A maioria dos dados sobre a infecção no Brasil, foi obtida a partir de amostras provenientes de bancos de sangue ou amostras de grupos específicos (gestantes, pacientes de clínicas de doenças sexualmente transmissíveis, coinfectados), que não necessariamente representam a população geral (Revisado por PROIETTI et al., 2005). Entretanto, o único estudo que demonstra dados com base populacional foi o realizado por Dourado e colaboradores (2003) na cidade de Salvador - Bahia, no qual estimou que cerca de 40 a 50 mil indivíduos estariam infectados com HTLV-1 em toda cidade.



**Figura 6.** Prevalência de HTLV-1/2 entre doadores de sangue em capitais de 26 estados brasileiros e no Distrito Federal (Adaptado de CATALAN-SOARES et al., 2005).

## 1.7 Epidemiologia molecular do HTLV-1

Considerando as diferenças nas sequências do gene *env* e da região *LTR* do DNA proviral, o HTLV-1 pode ser classificado em sete subtipos virais, sendo estes: “a” ou Cosmopolita (MIURA et al., 1994) que é encontrado em áreas endêmicas em todo o mundo; “b” ou Central Africano (HAHN et al., 1984; VANDAMME et al., 1994); “c” ou da Melanésia (GESSAIN et al., 1991); “d”, isolado de pigmeus em Camarões e no Gabão (CHEN et al., 1995; MAHIEUX et al., 1997), “e” no isolado de pigmeus na República Democrática do Congo (SALEMI et al., 1998); “f”, de um indivíduo do Gabão e “g”, recentemente descrito como um novo subtipo em Camarões, na África Central (WOLFE et al., 2005).

O subtipo Cosmopolita (a) é dividido em 6 subgrupos, a depender da sua origem e identificação geográfica dos primeiros isolados: A – Transcontinental, identificado em todo o mundo; B – Japonês, caracterizado em indivíduos de origem

japonesa; C – Oeste Africano, identificado em indivíduos oriundos de países do Oeste Africano; D – Norte Africano, identificado em indivíduos oriundos de países do Norte da África; E – Negros do Peru, identificado em indivíduos de cor negra no Peru e F – Etiópia, neste caso, o indivíduo era proveniente da Etiópia, na África (HAHN et al., 1984; GESSAIN et al., 1991; MIURA et al., 1994 e 1997; GASMI et al., 1994; VIDAL et al., 1994; VANDAMME et al., 1994; CHEN et al., 1995; SALEMI et al., 1998; VAN DOOREN et al., 1998; WOLFE et al., 2005; TREVIÑO et al., 2014) .

A distribuição dos subgrupos possui um determinado padrão, de acordo com a região ou regiões que foram inicialmente descritos. O subtipo Cosmopolita e subgrupo Transcontinental tem a maior distribuição mundial (GESSAIN e CASSAR, 2012), enquanto que os outros subgrupos, possuem algumas exceções, por exemplo: o HTLV-1aB foi reportado em japoneses e seus descendentes em outros lugares além do Japão, tais como Taiwan, Brasil, África do Sul e Peru (LIU et al., 1996; ZEHENDER et al., 2008; PESSÔA et al., 2014; BANDEIRA et al., 2015); o HTLV-1aC, além do Oeste da África, foi encontrado no Caribe e Guiana Francesa (ZEHENDER et al., 2008; DESRAMES et al., 2014) e o HTLV-1aD, oriundo do Norte da África, foi relatado em países do Oeste da África (DESRAMES et al., 2014).

## 1.8 Diagnóstico sorológico e molecular

O diagnóstico sorológico da infecção pelo HTLV-1/2 baseia-se na detecção de anticorpos específicos contra o vírus, através de testes sorológicos, como *ELISA* (*Enzyme-Linked Immunosorbent Assay*) e *Western Blot* (WB) (CONSTATINE et al., 1992). Estes podem ser classificados como testes de triagem e de confirmação.

O teste mais utilizado na triagem sorológica do HTLV-1/2 é o ensaio imunoenzimático (*Enzyme Immuno Assay*-EIA), no qual os antígenos específicos são adsorvidos a uma placa de poliestireno e a reação é revelada após a incubação do soro do indivíduo a um conjugado anti-IgG humano marcado. A reação é definida como positiva por intensidade colorimétrica, medida em densidade ótica (DO), a partir de um valor de corte definido ou *cut-off*.

Os ensaios de triagem detectam anticorpos contra o HTLV-1 e HTLV-2, porém não apresentam capacidade discriminatória entre essas infecções, fazendo-se necessária a confirmação do resultado por ensaios com alta especificidade e capazes

de distinguir a infecção causada pelo HTLV-1 daquela causada pelo HTLV-2, nestes casos, o ensaio mais utilizado é o teste WB (LAL et al., 1992).

O WB é capaz de distinguir a infecção pelo vírus tipo 1 ou 2, pois este teste possui antígenos específicos tanto para o HTLV-1 (MTA-1) quanto para o HTLV-2 (K-55). Adotou-se critérios que incluem a reatividade para as proteínas codificadas pelo gene *gag* (p19 e p24) e para as glicoproteínas do envelope viral (gp46 e gp21), nos casos das amostras consideradas positivas. Algumas amostras podem ter resultado positivo no WB, sem diferenciação entre os tipos 1 e 2 sendo consideradas positivas, porém não tipadas. Quando uma amostra reage com uma das bandas do WB, porém não completa o critério de positividade, seu resultado é considerado indeterminado.

Os resultados soro-indeterminados podem ser explicados pela imunossupressão em indivíduos coinfectados HTLV-1/HIV ou HTLV-1/2, além de soroconversão, especialmente em populações de risco ou em áreas endêmicas, ou por conta de infecções por cepas virais divergentes e/ou defectivas (THORSTENSSON et al., 2002).

As técnicas de biologia molecular, principalmente a reação em cadeia da polimerase (PCR), para diagnóstico confirmatório e diferencial da infecção pelos HTLV-1 e HTLV-2, fundamentam-se primariamente na detecção do ácido nucleico viral na forma de DNA proviral. Este diagnóstico tem contribuído na resolução dos padrões indeterminados do WB, na discriminação entre as infecções e coinfecção pelo HTLV-I e HTLV-II e no diagnóstico precoce da transmissão vertical.

Apesar da importância deste teste diagnóstico, ainda não existem testes comerciais para a realização da técnica da PCR para HTLV-1/2, assim, os laboratórios que realizam esse teste se baseiam em tecnologia própria, considerado *in house*. Além disso, os métodos moleculares ainda que apresentem alta sensibilidade e especificidade para o diagnóstico diferencial da infecção pelo HTLV-1/2, estão sujeitos a diversos fatores tais como: iniciadores inadequados; quantidade e qualidade do DNA alvo extraído; ausência de padronização da metodologia; resultados falso-positivos devidos à contaminação por produto de PCR e presença de inibidores de amplificação na amostra. Estes fatores são limitadores no diagnóstico correto da infecção pelo HTLV-1/2.

## 1.9 Vias de transmissão do HTLV-1

Há três vias associadas à transmissão do HTLV-1: a via vertical, sexual e parenteral. A via de transmissão vertical (mãe-filho), caracteriza-se pela transmissão transplacentária, infecção durante o parto e principalmente pelo aleitamento (CARNEIRO-PROIETTI et al., 2002), em que a eficácia da transmissão está relacionada a algumas variáveis, tais como: carga proviral e duração do aleitamento materno (BITTENCOURT, 1998; GONÇALVES et al., 2010).

A transmissão do HTLV-1, assim como de outras doenças sexualmente transmissíveis (DSTs) está associada ao sexo desprotegido, múltiplos parceiros sexuais e é claro, com o relacionamento com parceiro infectado pelo HTLV-1 (GONÇALVES et al., 2010). Estudos demonstram que a transmissão sexual do HTLV-1 é mais eficiente do homem para mulher, presume-se que isto seria consequência da maior quantidade de linfócitos infectados e viáveis liberados durante a ejaculação masculina (MURPHY, 1996; MANNIS et al., 1999). Entretanto, Roucoux e colaboradores (2005) observaram taxas semelhantes de transmissão sexual do HTLV-1 em ambos os sexos.

Finalmente, a via parenteral ocorre principalmente, por compartilhamento de agulhas/seringas contaminadas e por transfusão de sangue/hemoderivados, sendo esta forma a mais eficiente de transmissão, reforçando a importância da realização de testes de triagem sorológica nos serviços de hemoterapia (CARRAZZONE et al., 2004).

## 2 DOENÇAS ASSOCIADAS AO HTLV-1

Desde a sua identificação até os estudos mais atuais, o HTLV-1 tem sido relacionado à inúmeras manifestações clínicas que foram agrupadas de acordo com as suas características: (1) síndromes inflamatórias, tais como a TSP/HAM (Paraparesia Espástica Tropical/Mielopatia Associada ao HTLV-1) e síndrome do olho seco; (2) doenças malignas, representada pela ATL (Leucemia/Linforma de Células T do adulto) e finalmente, (3) complicações infecciosas, como a dermatite infecciosa.

### 2.1 Paraparesia Espástica Tropical/Mielopatia Associada ao HTLV-1 (TSP/HAM)

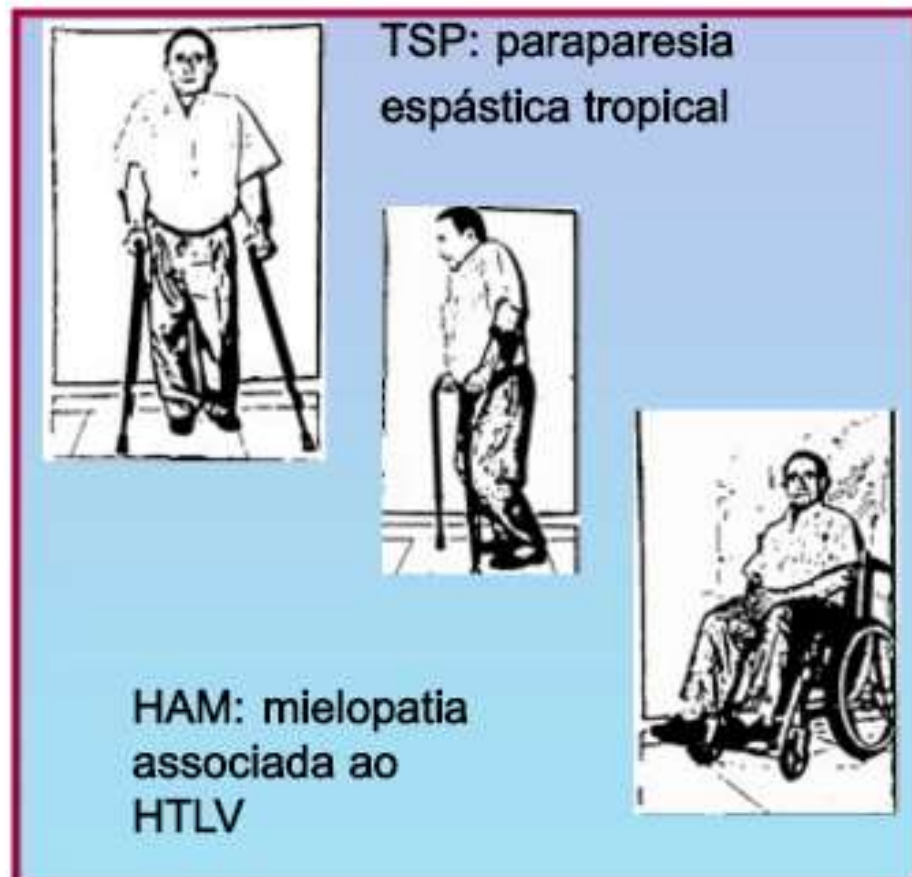
A TSP/HAM caracteriza-se como uma doença neurológica crônico-degenerativa, atingindo o sistema nervoso central (SNC), tendo como principal consequência a paraparesia dos membros inferiores (GESSAIN et al., 1985; OSAME et al., 1986) (Figura 7). Os critérios diagnósticos da TSP/HAM que são utilizados até os dias atuais, foram elaborados no final da década de 80 por Osame e colaboradores com o apoio da Organização Mundial da Saúde. Estes critérios baseiam-se principalmente nas manifestações neurológicas e no diagnóstico laboratorial (Quadro 1).

**Quadro 2.** Critérios de Diagnóstico de TSP/HAM (WHO, 1989).

<b>Principais Manifestações Neurológicas</b>
<ul style="list-style-type: none"> <li>• Paraparesia espástica crônica que progride geralmente de forma lenta, às vezes permanece inalterada após progressão inicial.</li> <li>• Fraqueza dos membros inferiores, de predomínio proximal.</li> <li>• Distúrbio vesical é uma característica precoce; constipação intestinal ocorre mais tardiamente; impotência e diminuição da libido são frequentes.</li> <li>• Sintomas sensitivos como formigamento, agulhadas e queimação, etc. são mais proeminentes do que sinais físicos objetivos.</li> <li>• Dor lombar baixa com irradiação para os membros inferiores é comum.</li> <li>• Sensibilidade vibratória é mais frequentemente comprometida que a proprioceptiva.</li> <li>• Hiperreflexia dos membros inferiores, frequentemente com clônus e sinal de Babinski.</li> <li>• Hiperreflexia dos membros superiores e os sinais de Hoffmann e de Trömner são frequentes; a fraqueza pode estar ausente.</li> <li>• Reflexo mandibular exaltado em alguns pacientes.</li> </ul>

<b>Diagnóstico Laboratorial</b>
<ul style="list-style-type: none"> <li>• Presença de anticorpos anti-HTLV-1 ou de antígenos no sangue e no LCR (líquido cefalorraquiano).</li> <li>• LCR pode apresentar pleocitose linfocitária moderada.</li> <li>• Linfócitos lobulados podem estar presentes no sangue e/ou no LCR.</li> <li>• Pode ocorrer hiperproteinorraquia leve a moderada.</li> <li>• Isolamento viral quando possível no sangue e/ou no LCR.</li> </ul>

(Adaptado do HTLV -Cadernos Hemominas, 2015).



**Figura 7.** Representação de um indivíduo com TSP/HAM.

Apesar dos inúmeros estudos relacionados à patogênese da TSP/HAM, os mecanismos foram pouco elucidados. Fatores tais como a carga proviral e a resposta imunológica do hospedeiro são associados à progressão da doença (BANGHAM e OSAME, 2005; LEPOUTRE et al., 2009). Esta patologia se desenvolve aproximadamente em 0.3-2% das pessoas infectadas (EDWARDS et al., 2011), afetam mais indivíduos do gênero feminino do que masculino, sendo que a maioria destes indivíduos tem o diagnóstico tardio da doença, cerca da 4<sup>a</sup> a 5<sup>a</sup> década de vida (GESSAIN et al., 1985; OSAME et al., 1986).



Com intuito de explicar a ação do HTLV-1 no desenvolvimento da TSP/HAM, três mecanismos principais foram postulados: (1) citotoxicidade direta, no qual considera-se que o dano neurológico está relacionado à ação antiviral dos linfócitos T citotóxicos (CTL). A infecção de astrócitos, neurônios e oligodendrócitos pelo HTLV-1, resultaria na apresentação de antígenos virais na superfície destas células, permitindo a ação desses linfócitos que agem liberando citocinas e assim, provocando a morte celular (IJICHI e OSAME, 1995; IJICHI et al., 1996; LEVIN; JACOBSON, 1997; NAKAMURA, 2000; FURUKAWA et al., 2003); (2) autoimunidade, uma resposta autoimune seria ativada pela reatividade cruzada de proteínas neuronais com anticorpos anti-Tax. Os linfócitos T CD4+ erroneamente gerariam uma resposta imunológica contra essas células (RODGERS-JOHNSON et al., 1990; GESSAIN e GOUT, 1992; FURUYA et al., 1998; LEE et al., 2005; MOSLEY et al., 2005); e por fim, (3) dano colateral, neste caso, os linfócitos T CD4+ infectados pelo HTLV-1 e os linfócitos T citotóxicos específicos contra o vírus migrariam através da barreira hematoencefálica, e no SNC causariam a degradação das células da glia, pela liberação de citocinas pró-inflamatórias, resultante da interação destes linfócitos (HÖLLSBERG; HAFLER, 1995; SAKAI et al., 2001; URETA-VIDAL et al., 2001; OSAME, 2002; KUBOTA et al., 2003; GOON et al., 2003; NAGAI et al., 2003; ARAÚJO et al., 2005; SILVA, 2006). Esta terceira hipótese é a mais aceita na atualidade, sendo atribuída como a principal causa do desenvolvimento de doença a resposta imune do hospedeiro (KANNAGI et al., 2004; KANNAGI et al., 2011; COOK et al., 2013).

A premissa que a grande maioria dos pacientes infectados com o HTLV-1 permaneça assintomática, traz o questionamento em relação à função da resposta imunológica na determinação do perfil clínico associado à infecção do vírus. Inúmeros fatores de interação entre vírus e o hospedeiro são possíveis candidatos para o aparecimento das patologias, embora os mecanismos ainda sejam uma incógnita (SABOURI et al., 2005). Estudos realizados por Bangham e Osame (2005) e Kashima e colaboradores (2006) não demonstraram associação entre as variantes do vírus e o risco para o desenvolvimento da TSP/HAM. Entretanto, outros estudos inferem que fatores tais como a resposta imunológica e características genéticas do hospedeiro estariam relacionadas ao desenvolvimento dos sintomas (CATALAN-SOARES et al., 2009; HADDAD et al., 2011). Um dos exemplos, seria o alelo HLA-DRB1\*0101 estar associado a uma maior susceptibilidade ao desenvolvimento da TSP/HAM, via ativação das células T CD4+ e o subsequente dano no SNC, do que os alelos HLA-A\*02 e HLA-Cw\*08, que ao contrário, estão relacionados à redução da carga proviral

e conseqüentemente, no menor risco para o desenvolvimento do TSP/HAM (BANGHAM, 2000; JEFFERY et al., 2000).

Diversos trabalhos associam uma alta carga proviral (CPV) na predição do desenvolvimento de TSP/HAM, apesar da existência de inúmeros pacientes que mesmo com CPV elevada, mantêm-se assintomáticos durante toda a vida (IWANAGA, 2010; GRASSI et al., 2013). Além disso, estudos recentes demonstram que níveis de CXCL10 (C-X-C motif chemokine 10), CXCL9 (C-X-C motif chemokine 9) e neopterin no líquido cefalorraquidiano (LCR) são candidatos à biomarcadores no prognóstico para a TSP/HAM, e a sua utilização na identificação precoce nos indivíduos infectados é importante para indicar uma terapia anti-inflamatória, possibilitando o abortamento das manifestações ligadas a esta mielopatia (SATO et al., 2013).

## 2.2 Leucemia/Linfoma de Células T do Adulto (ATL)

A ATL é considerada um linfoma/leucemia agressivo, sendo a primeira doença maligna identificada, causada por um retrovírus. Em relação ao seu curso clínico, é considerada uma doença heterogênea podendo ser classificada em 4 estágios: indolente ou *smouldering* (Figura 8), crônica, aguda e linfomatosa (Quadro 2) (SHIMOYAMA, 1991). Além disso, algumas manifestações clínicas restritas à pele, foram denominadas ATL cutânea e posteriormente subdivididas no tipo tumoral e eritematosa (BITTENCOURT et al., 2007). Cerca de 3-5% dos indivíduos infectados pelo HTLV-1, desenvolvem ATL (EDWARDS et al., 2011).

**Quadro 3.** Formas clínicas da ATL e suas principais características específicas.

Formas Clínicas da ATL	Características específicas
---------------------------	-----------------------------

<p><b>Forma indolente</b> (<i>Smouldering</i>)</p>	<ul style="list-style-type: none"> <li>• Presença de 5% ou mais de linfócitos T anormais no sangue periférico.</li> <li>• Ausência de linfocitose e de hipercalemia.</li> <li>• Desidrogenase láctica (LDH) aumentada até 1,5 vez o valor normal.</li> <li>• Ausência de linfadenopatia.</li> <li>• Derrames cavitários.</li> <li>• Envolvimento de fígado, baço, sistema nervoso central, ossos e trato gastrointestinal.</li> <li>• Pode haver comprometimento da pele e/ou dos pulmões.</li> </ul>
<p><b>Forma crônica</b></p>	<ul style="list-style-type: none"> <li>• Presença de linfocitose absoluta.</li> <li>• LDH até duas vezes o valor normal.</li> <li>• Ausência de hipercalemia.</li> <li>• Ascite ou derrame pleural.</li> <li>• Ausência de envolvimento do sistema nervoso central, ossos e trato gastrointestinal.</li> <li>• Pode apresentar envolvimento linfonodal, de fígado, baço, pele e/ou pulmões.</li> </ul>
<p><b>Forma linfomatosa</b></p>	<ul style="list-style-type: none"> <li>• Apresenta-se com linfadenomegalia,</li> <li>• Ausência de linfocitose e com 1% ou menos de linfócitos anormais no sangue periférico.</li> </ul>
<p><b>Forma aguda</b></p>	<ul style="list-style-type: none"> <li>• Presença de uma linfadenomegalia, hepatosplenomegalia, lesões cutâneas, ósseas, gastrointestinais e do sistema nervoso central.</li> <li>• Numerosos linfócitos atípicos no sangue,</li> <li>• É mais agressiva com uma evolução rápida para o óbito.</li> </ul>

(Adaptado de BITTENCOURT e FARRÉ, 2008).



**Figura 8.** Indivíduo com a forma clínica indolente da ATL apresentando lesões na pele. (Fonte: Bittencourt e FARRÉ, 2008)

A ATL é caracterizada pela infiltração de células T CD4+ malignas, que contém o DNA proviral do HTLV-1, nos linfonodos, baços, trato gastrointestinal e pele, além da presença de células T anormais com núcleo lobulado ou forma de flor (*flower cells*) (MATSUOKA, 2005).

A capacidade oncogênica do HTLV-1 está relacionada com a atividade reguladora de Tax, uma proteína que além de regular a expressão de genes virais, interage com fatores de transcrição celulares (CREB/ATF, NF- $\kappa$ B e SRF) e com moléculas de sinalização, gerando estímulos ou reprimindo a expressão de genes celulares. Além disso, estudos demonstram que esta proteína viral também induz o aumento da expressão de várias citocinas e seus receptores, envolvidos no crescimento e na proliferação de células T, fatores de transcrição, como HIF-1 (Fator de Indução de Hipóxia tipo 1), e proto-oncogenes. A proteína Tax ainda é capaz de reprimir a expressão ou inativar um conjunto de genes celulares que atuam como inibidores do crescimento celular, podendo inibir o reparo do DNA e os eventos de morte celular programada (MATSUOKA, 2005; YASUNAGA, 2007).

Os inúmeros estudos soroepidemiológicos indicam que a ATL e o HTLV-1 tem uma distribuição mundial. No entanto, a verdadeira prevalência de linfomas associados ao HTLV-1 é ainda largamente desconhecida (BLATTNER; POMBO DE OLIVEIRA, 1998).

As variantes genéticas virais (carga proviral, cepas virais, o baixo nível de anticorpos anti-Tax no soro), características do hospedeiro (polimorfismo de haplótipos HLA) ou do ambiente (modo de transmissão, dose do inóculo viral e idade da infecção) estão intimamente relacionados à patogênese da ATL e progressão da doença (HISADA, 1998; SLATTERY, 1999; PLANCOULAIN, 2000; YASHIK, 2001; BARMAK, 2003).

O sexo masculino apresenta uma maior chance (40% a mais) de desenvolver ATL em relação ao sexo feminino, apesar das mulheres terem uma maior prevalência de infecção pelo HTLV-1, justificado pela maior vulnerabilidade à transmissão do vírus por via sexual (GRANT, 2002). Outros fatores que predisõem à ATL, estão relacionados ao acometimento da dermatite infecciosa relacionada ao HTLV-1 durante a infância (HANCHARD, 1991) e um maior número de casos de ATL descritos em famílias, sugerindo uma predisposição genética para desenvolvimento deste tipo de leucemia (YASUNAGA, 2007).

No primeiro estágio da ATL, ocorre uma expressão transitória da proteína Tax em uma população limitada de células infectadas, deste modo, permite-se manter os níveis baixos desta proteína ao longo da infecção. Esta expressão de Tax tem um papel crítico na promoção da proliferação celular, na indução do ciclo celular, no acúmulo de mutações e anormalidades genéticas e na inibição da apoptose das células infectadas. Este cenário acontece no hospedeiro por um longo período e finalmente, a expansão clonal de populações celulares com DNA danificado levaria a um processo de transformação maligna de algum clone celular, desencadeando a ATL (YOSHIDA, 2001; 2010).

Do processo de infecção ao estabelecimento da doença, vários mecanismos de escape do sistema imunológico foram propostos: o primeiro estaria relacionado a baixa expressão da proteína viral Tax, controlada por proteínas virais, que atuam como reguladores negativos, exemplo a proteína HBZ. Sabe-se a proteína Tax atua na expansão de células infectadas e sua capacidade imunogênica, poderia tornar as células alvos fáceis do sistema imunológico do hospedeiro (SETO, 1991); deleções nas sequências dos genes *gag*, *pol*, *env* e na região *LTR* formando provírus defectivos (SHUH, 1999; TAKEDA, 2004); mutações nas sequências de Tax no local de ligação com o antígeno leucocitário humano HLA-A (FURUKAWA, 2001) e por fim, a ocorrência de modificações epigenéticas, tais como a hipermetilação do genoma celular, impedindo o processo de transcrição de genes (TANIGUCHI, 2005)

### 2.3 Dermatite Infecciosa associada ao HTLV-1

Apesar do HTLV-1 ter sido descoberto em 1980, a dermatite infecciosa já tinha sido descrita desde 1966 na Jamaica, sendo considerado uma forma grave de eczema infantil (SWEET, 1966). Mas somente anos mais tarde, em 1990, esta infecção foi relacionada ao HTLV-1 (LA GRENADÉ et al., 1990).

A dermatite infecciosa associada ao HTLV-1 é uma forma de eczema grave e recidivante, com lesões eritemato-descamativas, crostosas despontando-se em algumas partes do corpo ou de forma generalizada. Manifesta-se com maior frequência nas regiões retro-auriculares, pescoço, couro cabeludo, ombro (Figura 9) e genitália, normalmente associa-se à infecção bacteriana por *Staphylococcus aureus* e/ou *Streptococcus b-hemolyticus* (LA GRENADÉ et al., 1998). Estudos realizados por La Grenade e colaboradores (1996) conseguiram detectar em DNA de linfócitos extraídos de lesões de DIH e do sangue periférico, o genoma do HTLV-1, evidenciando a presença dos linfócitos infectados nas lesões.

Os mecanismos que levam ao desenvolvimento da DIH ainda não foram completamente elucidados, não se sabe por exemplo, o motivo no qual algumas crianças infectadas verticalmente desenvolvem a dermatite, enquanto que a maioria se mantém assintomática. Fatores como carga proviral elevada são indicativos de predisposição ao desenvolvimento da DIH (MALONEY et al., 2000).

Assim como na origem de outros processos inflamatórios não infecciosos, a interleucina 1 e 6 e o fator alfa de necrose tumoral, neste caso, originados pela ação transativadora da proteína Tax, estariam relacionados no desenvolvimento e manutenção das reações inflamatórias cutâneas, o que explicaria as constantes recidivas da DIH (TSCHACHLER et al., 1998).



**Figura 9.** Dermatite infecciosa em uma criança infectada pelo HTLV-I. Lesões eczematosas observadas predominantemente no couro cabeludo, face e região auricular. (Fonte: Bittencourt et al., 2005)

O HTLV-1 também é associado a outras doenças como a uveíte associada ao HTLV-1 (HAU), que se caracteriza por moderada a grave infiltração celular ocular e moderada vasculite retiniana (MOCHIZUKI et al., 1992; MOCHIZUKI et al., 1996); a síndrome seca, associada à xeroftalmia e a xerostomia; a artrite reumatoide, uma oligoartrite crônica de grandes articulações (NISHIOKA, 1996; MCCALLUM et al., 1997) e ao aumento da prevalência de algumas parasitoses, tais como a *estrongiloidíase* (PORTO et al., 2001).

Qualquer genoma está sujeito a mutações. Estas podem ser do tipo pontuais, quando ocorre uma troca de uma base por outra em um determinado sítio do DNA (ácido desoxirribonucleico), mutações do tipo inserções e deleções de um ou mais nucleotídeos (*Indels*).

A maioria das mutações ocorre em regiões promotoras e que não são influenciadas pela pressão seletiva, porém, quando ocorrem em regiões gênicas, podem gerar ou não alterações dos aminoácidos sintetizados e conseqüentemente da proteína. As mutações que não alteram a proteína são chamadas de sinônimas ou mutações silenciosas. Já aquelas que alteram, são as não-sinônimas e podem ser classificadas em dois tipos: mutações com sentido trocado ou *missense*, no qual ocorre a alteração do aminoácido e as mutações sem sentido ou *nonsense*, no qual resulta em um códon de parada (STENSON et al., 2008). Em relação às proteínas, a ocorrência de um *indel* quando múltiplo de três, resultará na inserção ou deleção do aminoácido, entretanto, quando não for múltiplo de três, ocasionará a mudança do quadro de leitura (*frameshift*), alterando a proteína.

Além de caracterizar diferentes subtipos do HTLV-1, as mutações pontuais e *indels* podem estar associadas à ocorrência de diferentes perfis clínicos em uma determinada infecção. Estas mutações podem modificar o genoma e conseqüentemente, alterar os seus produtos proteicos. Além disso, essas alterações no genoma são importantes para estudos que avaliam a origem e disseminação de uma determinada infecção (HARRIS et al., 2010).

No início deste projeto, apesar da importância clínica/epidemiológica do HTLV-1, existiam apenas 17 genomas completos do HTLV-1 que estavam disponíveis mundialmente (SEIKI et al., 1983; MALIK et al., 1988; EVANGELISTA et al., 1990; GESSAIN et al., 1993; CHOU et al., 1995; BAZARBACHI et al., 1995; ZHAO et al., 1993; WAZIRI et al., 1999; ELLERBROK et al., 1998; CAO et al., 2000; KASHIMA et al., 2004; OHSUGI et al., 2004; ANDONOV et al., 2012; ZANELLA L et al., 2012), destes, somente sete sequências possuíam informação em relação ao perfil clínico do paciente infectado (quatro sequências de indivíduos com ATL, duas sequências de indivíduos com TSP/HAM e uma sequência de indivíduo assintomático). Além disso, apenas duas sequências possuíam informação sobre a data de coleta da amostra, dado importante para análises evolutivas de coalescência e finalmente, nenhuma sequência possui informação sobre a carga proviral, importante marcador para progressão da TSP/HAM, por exemplo (MONTANHEIRO et al., 2005).

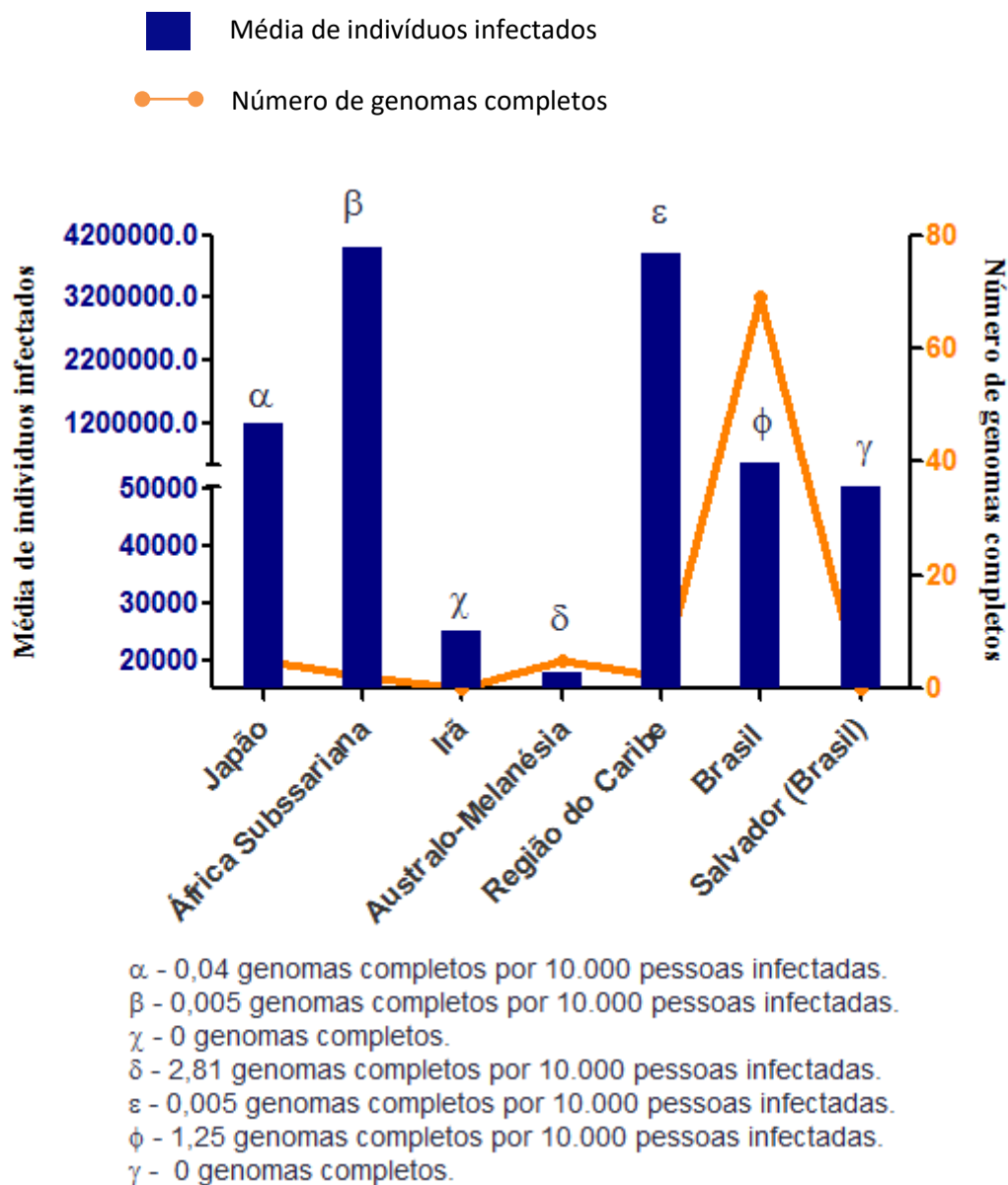


Nos anos posteriores, novos genomas completos foram gerados (CASSAR et al., 2013; MA et al., 2013; PÊSSOA et al., 2014; KURAMITSU M et al., 2015; ZANELLA et al., 2016; ENOSE-AKAHATA et al., 2016), atualmente, cerca de 91 deles estão disponíveis publicamente. As sequências e informações citadas acima podem ser identificadas no Banco de Dados intitulado “*HTLV-1 Molecular Epidemiology Database*” (<http://htlv1db.bahia.fiocruz.br/>), desenvolvido pelo nosso grupo de pesquisa, com colaboração do *Rega Institute (Katholieke Universiteit Leuven, Bélgica)*.

A maioria dos estudos que avaliaram a soroprevalência do HTLV-1 foram realizados em doadores de sangue, mulheres grávidas ou pacientes hospitalizados. Entretanto, os dados mais concretos são os gerados a partir de estudos que avaliam a população geral, similar ao realizado por Dourado e colaboradores na cidade de Salvador (HLELLA et al., 2009).

Há uma grande dificuldade em desenhar o cenário real da infecção pelo HTLV-1 em todo o mundo. Em 2012, a partir de um artigo de Gessain e Cassar no qual revisou-se os dados de soroprevalência do HTLV-1 publicados mundialmente, foi possível estimar um número provável de indivíduos infectados em todo mundo. A partir desses dados, realizou-se uma avaliação comparativa entre a média de indivíduos infectados no mundo e o número de genomas completos, encontramos cerca de 0,12 genomas completos por 10.000 indivíduos infectados no mundo. Considerando as regiões endêmicas para HTLV-1 versus o número de genomas completos gerados a partir de indivíduos dessas regiões, o Brasil obteve a segunda melhor média, com 1,25 genomas completos por 10.000 indivíduos infectados (Figura 10), entretanto, todos estes genomas completos são provenientes de indivíduos residentes em São Paulo.

Levando-se em consideração o cenário da epidemiologia do HTLV-1 no Brasil, a cidade de Salvador se destaca por apresentar a maior soroprevalência do vírus, ainda assim, não há genomas completos do HTLV-1 caracterizados de indivíduos oriundos dessa região. Deste modo, este trabalho contribuiu na geração de genomas virais nessa população endêmica para a infecção pelo HTLV-1, e sobretudo, na produção de sequências genômicas bem anotadas e caracterizadas, com informações clínicas e epidemiológicas dos hospedeiros. À vista disso, esses genomas poderão ser bem explorados para investigações moleculares sobre a infecção pelo HTLV-1.



**Figura 10.** Análise comparativa entre a média de indivíduos infectados pelo HTLV-1 nas regiões endêmicas versus o número de genomas completos disponíveis.

As novas tecnologias de sequenciamento, denominadas de alta performance ou alta demanda promovem o sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões de pares de bases em uma única corrida. Em relação ao sequenciamento de Sanger, estas novas plataformas geram um maior número de dados, com uma grande economia de tempo e custo por base sequenciada (MARDIS, 2008).

Especificamente nos estudos em retrovirologia, estes novos sequenciamentos surgem da necessidade de entender a evolução, funcionamento, epidemiologia

molecular e variabilidade destes vírus (RADFORD et al., 2012; CAPOBIANCHI, 2013; GIOMBINI, 2013; ROZERA, 2013). Estas novas plataformas possibilitam estudos mais eficientes dos genomas, assim como, das mutações que diferenciam estes genomas (HOLT et al., 2008; BROWN, HARRIS et al., 2010; FISHWICK; CHOKSHI, 2011).

O sequenciamento de novos genomas completos, a disponibilidade de informações tais como perfil clínico, carga proviral e data de coleta da amostra, adicionado à caracterização genética destes genomas poderá favorecer aos estudos relacionados à patogenicidade do HTLV-1 e na compreensão da sua história natural e adaptação em hospedeiros humanos.

## 4 OBJETIVOS

### 4.1 Geral

Caracterizar genomas completos e parciais do vírus linfotrópico de células T humanas do tipo 1 (HTLV-1).

### 4.2 Específicos

- Analisar as variáveis moleculares, clínicas e epidemiológicas dos dados depositados no banco de dados previamente publicado (*HTLV-1 Molecular Epidemiology Database*).
- Sistematizar a mineração de dados e subtipagem do HTLV-1;
- Gerar e disponibilizar genomas completos provenientes de pacientes infectados pelo HTLV-1 com diferentes perfis clínicos: indivíduos com TSP/HAM, ATL, DIH e assintomático;
- Identificar as possíveis variantes nos genomas completos provirais do HTLV-1 que estejam associados aos diferentes perfis clínicos: indivíduos com TSP/HAM, ATL, dermatite infecciosa e assintomático;
- Calcular a variabilidade genética no genoma completo proviral do HTLV-1 em cada grupo de estudo: indivíduos com TSP/HAM, ATL, DIH e assintomático;
- Caracterizar os sítios de ligação para fatores de transcrição nas sequências promotoras (LTR) do HTLV-1 nos grupos: indivíduos com TSP/HAM, ATL, DIH e assintomático;
- Avaliar o impacto da variabilidade genética na região LTR em relação à afinidade de ligação dos fatores de transcrição aos sítios correspondentes;
- Analisar a relação entre a variabilidade genética nas regiões codificantes, em cada grupo de estudo, e a possibilidade de perda ou aquisição de diferentes sítios de modificações pós-traducionais.

## 5 RESULTADOS

Todos os resultados obtidos foram organizados nos seguintes formatos: 1- Artigo publicado na revista Memórias do Instituto Oswaldo Cruz em 2014; 2- Capítulo do livro: *Human T-Lymphotropic Viruses - Methods and Protocols'* editora Springer (2016) e 3 – Artigo 2: manuscrito em preparação.

### **5.1 Artigo I: Inferências sobre o cenário global da infecção pelo vírus linfotrópico de células T humanas do tipo 1, utilizando data mining de sequências virais.**

Neste artigo intitulado “**Inferences about the global scenario of human T-cell lymphotropic virus type 1 infection using data mining of viral sequences**”, publicado na revista Memórias do Instituto Oswaldo Cruz em 2014, objetivou-se descrever os cenários clínicos, moleculares e epidemiológicos da infecção pelo HTLV-1 através das sequências armazenadas no banco de dados secundário *HTLV-1 Molecular Epidemiology Database* (ARAUJO et al, 2012). Todas as análises descritivas foram realizadas a partir do algoritmo de busca implementado no próprio banco de dados, assim, de acordo com as variáveis selecionadas, os resultados eram gerados e tabulados no formato de arquivo *xls* (*Excel – Microsoft Office*).

Um total de 2.545 sequências completas e parciais do HTLV-1 foram registradas e destas, 1.967 (77.3%) representavam isolados únicos. Entre esses isolados, 93% possuíam informações de origem geográfica e apenas 39% estavam relacionados a qualquer perfil clínico. Um total de 1.091 sequências tinham informações sobre a origem geográfica e o subtipo viral concomitantemente, sendo que a grande maioria (93%), foram identificadas como subtipo Cosmopolita (HTLV-1a).

Os dados étnicos, idade e gênero dos indivíduos dos quais as sequências foram geradas, foram muito escassos. Das sequências analisadas, 797 tinham informações sobre o perfil clínico (TSP / HAM, 43%; ATL, 19 %; assintomático, 32,39% e outras doenças, 5,61%).

Embora a mineração de dados tenha permitido inferências sobre aspectos específicos da infecção pelo HTLV-1, devido à relativa escassez de dados das sequências disponíveis, não foi possível delinear um cenário global de infecção pelo HTLV-1. Ressalva-se a importância dos dados clínicos e epidemiológicos dos indivíduos, dos quais as sequências foram provenientes para implantação de políticas públicas de saúde em todo o mundo.

## **Inferences about the global scenario of human T-cell lymphotropic virus type 1 infection using data mining of viral sequences**

Thessika Hiaila Almeida Araujo<sup>1</sup>, Fernanda Khouri Barreto<sup>1</sup>, Luiz Carlos Júnior Alcântara<sup>1,+</sup>,  
Aline Cristina Andrade Mota Miranda<sup>2</sup>

1 Centro de Pesquisa Gonçalo Moniz-Fiocruz, Salvador, BA, Brasil

2 Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, BA, Brasil

### **ABSTRACT**

Human T-cell lymphotropic virus type 1 (HTLV-1) is mainly associated with two diseases: tropical spastic paraparesis/ HTLV-1-associated myelopathy (TSP/HAM) and adult T-cell leukaemia/lymphoma. This retrovirus infects five-10 million individuals throughout the world. Previously, we developed a database that annotates sequence data from GenBank and the present study aimed to describe the clinical, molecular and epidemiological scenarios of HTLV-1 infection through the stored sequences in this database. A total of 2,545 registered complete and partial sequences of HTLV-1 were collected and 1,967 (77.3%) of those sequences represented unique isolates. Among these isolates, 93% contained geographic origin information and only 39% were related to any clinical status. A total of 1,091 sequences contained information about the geographic origin and viral subtype and 93% of these sequences were identified as subtype "a". Ethnicity data are very scarce. Regarding clinical status data, 29% of the sequences were generated from TSP/HAM and 67.8% from healthy carrier individuals. Although the data mining enabled some inferences about specific aspects of HTLV-1 infection to be made, due to the relative scarcity of data of available sequences, it was not possible to delineate a global scenario of HTLV-1 infection.

Key words: HTLV-1 - data mining - HTLV-1 database

Human T-cell lymphotropic virus type 1 (HTLV-1) was the first described human retrovirus (Poiesz et al. 1980). This retrovirus is the causative agent of tropical spastic paraparesis/HTLV-1-associated myelopathy (TSP/HAM) (Gessain et al. 1985), adult T-cell leukaemia/lymphoma (ATL) (Yoshida et al. 1982) and other inflammatory diseases such as HTLV-1-associated infectious dermatitis (La Grenade et al. 1998) and HTLV-1-associated uveitis (Mochizuki et al. 1992). However, the pathogenesis of some clinical manifestations is not yet fully understood.

Epidemiological data show that HTLV-1 has a worldwide distribution and it is estimated that five-10 million people are infected (Gessain & Cassar 2012). This infection is endemic in southwestern Japan (Mueller et al. 1996), sub-Saharan Africa (Gessain & de Thé 1996), regions of the Caribbean (Hanchard et al. 1990) and minor areas in Iran, Melanesia (Mueller 1991) and Brazil (Galvão-Castro et al. 1997).

Regardless, HTLV-1 epidemiology still presents many challenges. Virus prevalence rates have been correlated with geographic characteristics and the social setting of destitute populations. However, these populations are not frequently the target of great public and government interest (Galvão-Castro et al. 1997). Molecular studies, especially during the late decade, have contributed to the acquisition of knowledge about virus epidemiology and the molecular characteristics. Furthermore, a great amount of viral sequences are generated from these molecular studies because of this appropriate data management and data mining can provide additional consistent information about HTLV-1 infection.

In response to the need of obtaining more information about the already generated and available HTLV-1 sequences, HTLV-1 Molecular Epidemiology Database ([htlv1db.fiocruz.bahia.br](http://htlv1db.fiocruz.bahia.br)) was developed (Araujo et al. 2012). This database contains information that can support our understanding of viral pathogenesis, the route of transmission, polymorphisms, epidemiology, genotype-phenotype relationships, geographic distribution and viral evolution. Therefore, the purpose of the present study was to assess the different types of information deposited in HTLV-1 Molecular Epidemiology Database to describe clinical, molecular and epidemiological scenarios about HTLV-1 infection.



## MATERIALS AND METHODS

This is a descriptive study about the clinical, molecular and epidemiological data of HTLV-1 infection that are associated with the stored genetic sequences in HTLV-1 Molecular Epidemiology Database.

All the descriptive analyses were performed using the search algorithm implemented at the HTLV-1 database (Araujo et al. 2012). Initially, we made a list with the variables (age, gender, clinical status, subtype, subgroup, geographic origin) that were more frequent in the database. We then performed combinations with the listed variables; for example, we searched for sequences with information about geographic origin, viral subtype and clinical status. These combinations constitute the section subheadings of the Results and Discussion section. The HTLV-1 database allows all the search results to be organised as spread sheets for further analyses.

After the search step and the generation of spread sheets containing the results, we used the Excel program to perform descriptive analyses.

## RESULTS AND DISCUSSION

Currently, HTLV-1 Molecular Epidemiology Database stores 2,545 HTLV-1 sequences, 1,967 (77.3%) of which represent different isolates. These 1,967 sequences were selected for this study and 91 (3.6%) other distinct sequences, which did not have information about the viral isolate, were also included. Ultimately, 2,058 sequences and their data were analysed.

Geographic origin, viral subtype and clinical status among viral sequences - Among the 2,058 viral sequences, 1,914 (93%) were associated with geographic origin in the GenBank notes. Fig. 1 shows the distribution of HTLV-1 sequences among different geographic regions: 1.6% of the sequences originated from HTLV-1 isolates from North America, 2.4% from Oceania, 3% from Europe, 3.3% from Central America, 17.7% from Africa, 32% from Asia and 40% from South America. With regard to the South America sequences, most of the isolates were from HTLV-1 infections in Brazil (55%) and Argentina (22.1%), as shown in Fig. 2.

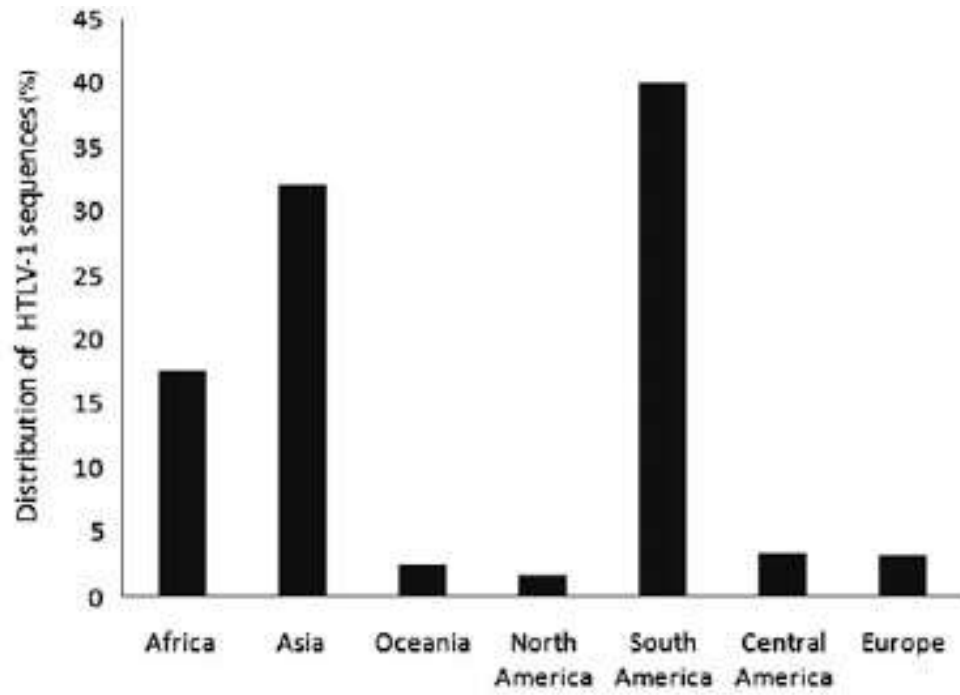


Fig. 1: geographical distribution of 2,058 sequences stored in the HTLV-1 Molecular Epidemiology Database.



Fig. 2: distribution (%) of human T-cell lymphotropic virus type 1 sequences among the countries in South America.

Although HTLV-1 infection has a cosmopolitan geographic distribution, it has a heterogeneous distribution, such that Asia and South America are characterised as endemic areas (Proietti et al. 2005, Carneiro-Proietti et al. 2006). This heterogeneous distribution is also represented in the distribution of sequences available in GenBank and in the number of exploratory studies of HTLV-1 infection developed in each geographical region.

Several studies have reported a high prevalence of HTLV infection in Africa (Proietti et al. 2005); however, there are few sequences about this geographic region deposited in GenBank. This profile is frequent and emphasise that it is necessary to increase the use of molecular data as an important tool of epidemiology investigation. This result suggests that it is necessary to create new health units that could be able to perform molecular diagnosis and, therefore, could generate and record molecular data about new HTLV-1 cases. Nonetheless, the high prevalence of HTLV infection in Asia and parts of South America (Carneiro-Proietti et al. 2006, Sonoda et al. 2011) corroborates the high amount of sequences assembled in the database. However, it is possible to observe a lack of clinical and epidemiological information in the GenBank annotations. This observation shows that the authors should provide the maximum amount of information as possible because the molecular data could be useful for many different inferences about HTLV-1 infection and, therefore, useful for encouraging the politics of prevention.

The search about the number of HTLV-1 sequences with geographic origin and viral subtype showed that 1,091 contained information for both variables. The results showed that 1,019 (93.4%) sequences were classified as subtype "a" and that this subtype had a worldwide distribution among the sequences deposited in the database. This higher prevalence can be attributed to the fact that it is the worldwide subtype found especially in Japan, the Caribbean, South America and Africa. Subtypes "b" (4.9%), "c" (0.5%), "d" (0.6%), "e" (0.1%), "f" (0.2%) and "g" (0.2%) were distributed in specific regions (Fig. 3). These subtypes are usually restricted to certain areas, such as subtype "c", found in Australia-Melanesia (Galvão-Castro et al. 1997).

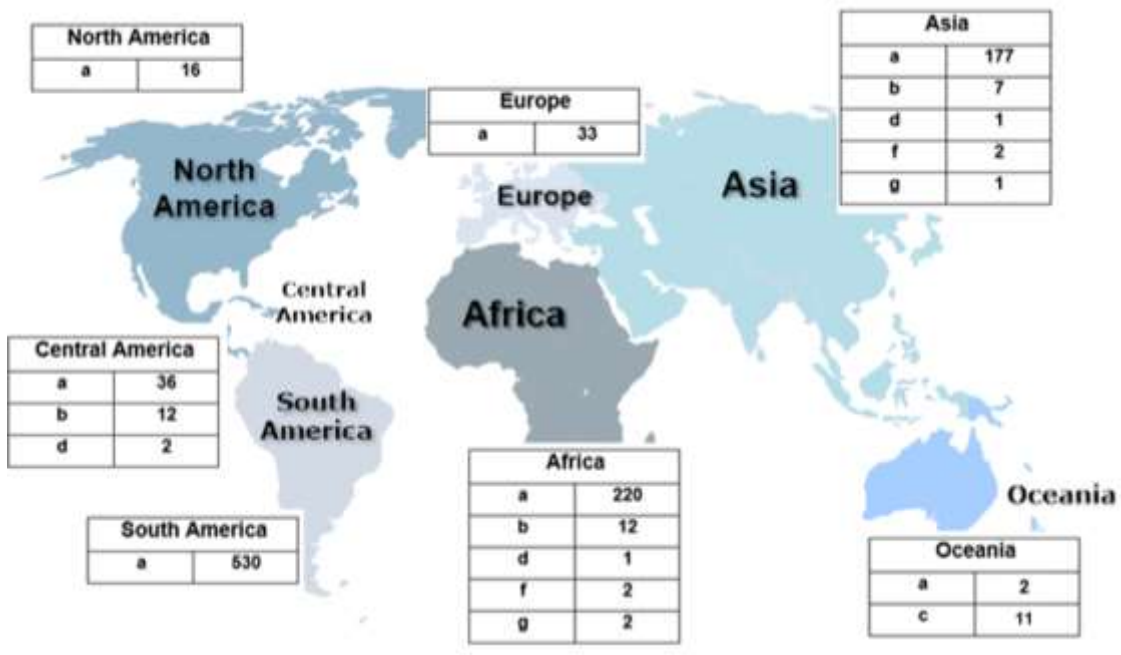


Fig. 3: geographical distribution of absolute number of human T-cell lymphotropic virus type 1 (HTLV-1) sequences stored in the HTLV-1 Molecular Epidemiology Database, using the subtype variable

Finally, using the geographic origin and viral subtype, we performed a search for clinical status. It was possible to identify that 279 sequences had information for these three variables in the GenBank annotations. Regarding these sequences, 35.8% originated from Asia, 32.2% from South America, 14.3% from Africa, 10% from Central America, 2.8% from Europe, 4.3% from North America and 0.3% from Oceania. Our analyses showed that 86.3% sequences were identified as subtype "a", 12.5% sequences were subtype "b" and 1.1% were subtype "d". Concerning the clinical status, 158 sequences (subtypes "a", 88.6%; "b", 10.7%; "d", 0.6%) were derived from HTLV-1 infection in healthy carrier (HC) individuals, 71 sequences (subtypes "a", 80.2%; "b", 18.3%; "c", 1.5%) were from TSP/HAM individuals and 34 sequences (subtypes "a", 91.2%; "b", 8.8%) were from ATL individuals; 19 viral sequences (subtypes "a", 84.2%; "b", 10.5%; "d", 5.3%) were related to other diseases not yet described.

Clinical status and viral subtype and subgroup of viral sequences from South America - Using HTLV-1 Molecular Epidemiology Database enabled the observation of the overall global scenario of HTLV-1 sequences from South America. From all the collected sequences from South America, 77 were from HC individuals, 63 from TSP/HAM individuals and 12 from ATL individuals; six viral sequences from individuals with other diseases. However, 609 sequences did not have any information about the clinical status in the GenBank annotations. As most of the studies of HTLV-1 infection in South America have been developed in Argentina and Brazil, the greatest number of sequences and, therefore, the greatest number of molecular and epidemiological data are also generated from these regions. Because of this fact, it is

important to emphasise the need of developing new exploratory studies about HTLV-1 infection in other countries in South America.

Only 77 of the sequences from HTLV-1 cases of infection in South America had information about the clinical status, viral subtype and subgroup, at the same time, in the GenBank annotation. Among these sequences, 18 were from TSP/HAM individuals, four sequences were from ATL individuals and six sequences were from individuals with other HTLV-associated diseases. The greatest number of sequences was generated from HC individuals (n = 49). All of the 77 sequences were identified, in the GenBank annotation, as subtype "a", which was the subtype most found among infected individuals in South America. With regard to the subgroup classification, 70 sequences were identified as subgroup "a" (n = 48 HC, n = 2 ATL, n = 15 TSP/HAM, n = 5 other diseases), three sequences as subgroup "b" (n = 2 TSP/HAM, n = 1 ATL) and three sequences as subgroup "c" (n = 1 ATL, n = 1 HC, n = 1 TSP/HAM); only one sequence was identified as subgroup "e" (lymphoma).

Clinical status, age, gender, viral subtype and ethnicity among viral sequences - The investigation about clinical status separately showed that 797 of the stored sequences were related to one HTLV-1-associated clinical status (TSP/HAM, 43%; ATL, 19%; HC, 32.39%; other diseases, 5.61%). The other HTLV-1-associated diseases, such as dermatitis and sicca syndrome, were not reported in any annotation of the stored viral sequences. Using the information about the clinical status, we searched for additional data such as gender, age, viral subtype and ethnicity (Table). Approximately 15.2% of the sequences contained information about the infected patient's gender and 10.8% of the sequences provided the age of the infected patient.

The data about ethnicity were very scarce, as only 41 (5%) of the 797 stored sequences had information about ethnic origin in the GenBank annotation. Approximately 78% of the sequences (n = 41) had information about gender, clinical status, viral subtype, ethnicity and age at the same time. All originated from women infected by one HTLV-1 subtype "a" isolate and 27.5% of these women were younger than 40 years old. Regarding clinical status, 29% of the sequences originated from TSP/HAM women and 67.8% sequences from HC women; 3.2% sequences were generated from infected women with other HTLV-1-associated diseases.

TABLE 1 Distribution of gender, subtype and geographic origin among the clinical status: tropical spastic paraparesis/human T-cell lymphotropic virus type 1 (HTLV-1)-associated myelopathy (TSP/HAM), adult T-cell leukaemia/lymphoma (ATL), healthy carrier (HC) and other HTLV-1 associated disease.

	TSP/HAM	ATL	HC	Others <sup>a</sup>
	n = 343	n = 137	n = 261	n = 45
Clinical status	n (%)	n (%)	n (%)	n (%)
<hr/>				
Sequences with information about sex	40 (11.6)	13 (9.4)	89 (34)	16 (35.5)
Male	15 (37.5)	6 (46.2)	33 (37)	8 (50)
Female	25 (62.5)	7 (53.8)	56 (63)	8 (50)
Sequences with information about subtype	71 (20.7)	36 (26.3)	158 (60.5)	19 (42.2)
Subtype				
“a”	57 (80.2)	33 (91.6)	140 (88.6)	16 (84.2)
“b”	13 (18.4)	3 (8.4)	17 (10.8)	2 (10.5)
Other	1 (1.4)	0 (0)	1 (0.6)	1 (5.3)
Sequences with information about geographic origin	320 (93.4)	128 (93.4)	260 (99.6)	45 (100)

a: sequences from patients with either infective dermatitis, histoplasmosis, stroke, seborrhoea dermatitis, leprosy, nonspastic paraparesis, facial palsy or leg paresis.

A great number of studies show that some ethnically defined factors are likely to be associated with HTLV-1 persistence and the development of ATL or TSP/HAM among HTLV-1 endemic populations. Therefore, this information should be further investigated in HTLV-1 infection cases. Furthermore, new studies about the genetic background of infected individuals by the analysing polymorphic determinants of human leukocyte antigen alleles and their immune responsiveness to HTLV-1 are important points in the approach of the ethnic factors involved in HTLV-1 clustering and the disease segregation of ATL and TSP/HAM (Sonoda et al. 2011).

HTLV-1 Molecular Epidemiology Database enabled some inferences about the specific aspects of HTLV-1 infection. However, due to the relative scarcity of data about the available sequences, it was not possible to delineate a global scenario of HTLV-1 infection. Molecular and epidemiological data for viral sequences should be offered more frequently because this information can be used for planning public health policies.

## REFERENCES

Araujo THA, Souza-Brito LI, Libin P, Deforche K, Edwards D, Albuquerque-Junior AE, Vandamme AM, Galvão-Castro B, Alcântara LCJ 2012. A public HTLV-1 molecular epidemiology database for sequence management and data mining. PLoS ONE 7: e42123.

Carneiro-Proietti ABF, Catalan-Soares BC, Castro-Costa CM, Murphy EL, Sabino EC, Hisada M, Galvão-Castro B, Alcântara LCJ, Remondegui C, Verdonck K, Proietti FA 2006. HTLV in the Americas: challenges and perspectives. *Rev Panam Salud Publica* 19: 44-53.

Galvão-Castro B, Loures L, Rodrigues LG, Sereno A, Ferreira OC, Franco LGP, Muller M, Sampaio DA, Santana A, Passos LM, Proietti F 1997. Distribution of human T-lymphotropic virus type I among blood donors: a nationwide Brazilian study. *Transfusion* 37: 242-243.

Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de Thé G 1985. Antibodies to human T lymphotropic virus type I in patients with tropical spastic paraparesis. *Lancet* 2: 407-409.

Gessain A, Cassar O 2012. Epidemiological aspects and world distribution of HTLV-1 infection. *Front Microbiol* 3: 388.

Gessain A, de Thé G 1996. Geographic and molecular epidemiology of primate T lymphotropic retroviruses: HTLV-I, HTLV-II, STLV-I, STLVP and PTLV-L. *Adv Virus Res* 47: 377-426.

Hanchard B, Gibbs WN, Lofters W, Campbell M, Williams E, Williams N, Jaffe E, Cranston B, Panchoosingh LD, Blattner WA, Manns A 1990. *Human retrovirology: HTLV*, Raven Press, New York, p. 173-183.

La Grenade L, Manns A, Fletcher V, Derm D, Carberry C, Hanchard B, Maloney EM, Cranston B, Williams NP, Wilks R, Kang EC, Blattner WA 1998. Clinical, pathologic and immunologic features of human T lymphotropic virus type I-associated infective dermatitis in children. *Arch Dermatol* 134: 439-444.

Mochizuki M, Watanabe T, Yamaguchi K, Tajima K, Yoshimura K, Nakashima S, Shirao M, Araki S, Miyata N, Mori S 1992. Uveitis associated with human T lymphotropic virus type I: seroepidemiologic, clinical and virologic studies. *J Infect Dis* 166: 943-944.

Mueller N 1991. The epidemiology of HTLV-1 infection. *Cancer Causes Control* 2: 37-52.

Mueller N, Okayama A, Stuver S, Tachibana N 1996. Findings from the Miyazaki Cohort Study. *J Acquir Immune Defic Syndr Hum Retrovirol* 13 (Suppl. 1): S2-S7.

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci USA* 77: 7415-7419.

Proietti FA, Carneiro-Proietti ABF, Catalan-Soares BC, Murphy EL 2005. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* 24: 6058-6068.

Sonoda S, Li HC, Tajima K 2011. Ethnoepidemiology of HTLV-1 related diseases: ethnic determinants of HTLV-1 susceptibility and its worldwide dispersal. *Cancer Sci* 102: 295-301.

Yoshida M, Miyoshi I, Hinuma Y 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc Natl Acad Sci USA 79: 2031-2035.

## 5.2 Capítulo de livro: Banco de dados, *data mining* e genotipagem do HTLV-1

O capítulo do livro intitulado “***Molecular epidemiology database for HTLV-1 sequence management, data mining and genotyping***” compreende um protocolo que ensina de forma didática os pesquisadores ou interessados na área, a utilizar a ferramenta de Subtipagem do HTLV-1 (ALCANTARA et al., 2009), que consiste em um *software on-line* para a genotipagem do HTLV-1 e o banco de dados sobre a Epidemiologia Molecular do HTLV-1 (ARAUJO et al., 2012), uma ferramenta que



permite ao usuário o gerenciamento de sequências, a mineração de dados, além de possibilitar o *download* dos dados selecionados: sequências no formato *fasta* e informações clínicas e demográficas tabuladas em um arquivo *xls* do programa Excel (*Microsoft Office*). Estes dois programas *online*, propiciam estudos relacionados à epidemiologia molecular e genotipagem do HTLV-1. O capítulo já foi enviado para publicação no livro intitulado “*Human T-Lymphotropic Viruses - Methods and Protocols*” editora *Springer* (2016).

**Molecular epidemiology database for HTLV-1 sequence management, data mining and genotyping.**

Thessika Hialla Almeida Araújo, Filipe Ferreira de Almeida Rego and Luiz Carlos Junior Alcantara

Affiliations

Thessika Hialla Almeida Araújo, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil.

Filipe Ferreira de Almeida Rego, Fundação Oswaldo Cruz (FIOCRUZ) and Universidade Católica do Salvador, Salvador, Bahia, Brazil.

Luiz Carlos Junior Alcantara, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil.

Corresponding author:

Luiz Carlos Junior Alcantara, Ph.D.

FIOCRUZ

Rua Waldemar Falcão 121, Candeal, Salvador, Bahia, Brazil 40295-001.

Telephone: +55 (71) 3176-2255

Fax: +55 (71) 3176-2300

email: [lalcan@bahia.fiocruz.br](mailto:lalcan@bahia.fiocruz.br)

## Summary

A central database to aggregate sequence information from a range of epidemiological aspects including HTLV-1 pathogenesis, origin, and evolutionary dynamic would be useful to scientists and physicians worldwide. This Chapter describes two online tools for studies related to HTLV-1, The HTLV-1 Molecular Epidemiology Database and the HTLV-1 Subtyping Tool. The HTLV-1 Molecular Epidemiology Database is a tool for sequence management and data mining which allows researchers to download sequences with clinical and demographic information. The HTLV-1 Subtyping Tool is an online software used for HTLV-1 genotyping, the algorithm consists in the alignment of a query sequence with a carefully selected set of predefined reference strains, followed by phylogenetic analysis.

## Key Words

HTLV-1 genotyping tool, HTLV molecular epidemiology, HTLV-1 database

## Introduction

The sequence analysis of HTLV is being a very useful tool with different approaches depending on the analyzed genomic region. The analysis of the LTR and envelope sequences of this virus is being used, in addition to mtDNA and Y chromosome analysis, to study the human migration and virus genotyping [1]. The HTLV pX gene analysis are usefull to understand the functions of its generated proteins and their location into the cell [2] while the analysis of env, pol and gag sequences has improved the knowledge of possible epitopes for vaccine design and helping to improve diagnostic methods [3, 4].

The HTLV nucleotide sequences, as well as from other viruses, are usually stored in the GenBank (<http://www.ncbi.nlm.nih.gov>). Despite the importance of analyze HTLV sequences, the GenBank, a primary database, does not have the task of giving appropriated sequences with clinical and demographic data for analysis and does not genotype viruses. Because of this was developed the HTLV-1 Molecular Epidemiology Database, an online data

mining tool that retrieves and stores annotated HTLV-1 proviral sequences from clinical, epidemiological, and geographical studies and the HTLV-1 Subtyping Tool [5–7].

## Materials

In this case, all is needed is a computer with Internet connection and a spreadsheet reader. For analysis beyond the scope of this section, most of the software available can be viewed at <http://www.bioinformaticssoftwareandtools.co.in/index.php>.

## Methods

### HTLV-1 Molecular Epidemiology Database

The website interfaces were developed in HTML and sever-side scripting written in PHP. The website interface contains specific search fields to allow various data combinations. User queries create a form (form tag) containing the values (variables) selected. This form generates a script that retrieves the data stored in the MySQL database. A second script organizes the data for display on the website, allowing for visualization of the information with the option to download the organized data. The developed database provides information regarding the indexed sequences in GenBank. In addition, all the sequences were genotyped using the LASP HTLV-1 Automated Subtyping Tool [7]. The user is able to choose search criteria and perform a query to generate an output of relevant sequences and information. The sequence output may be downloaded in FASTA format and the information table in Microsoft Excel spreadsheets .xls format. The HTLV-1 Molecular Epidemiology database is hosted on the Gonçalo Moniz Research Center/Oswaldo Cruz Foundation Research Center server with access at <http://htlv1db.bahia.fiocruz.br/>.

### Downloading sequences

1. The HTLV-1 Molecular Epidemiology Database homepage displays interface (figure 1) that contains numerous fields for refining database queries;
2. One or more fields may be selected;
3. The remaining unused fields will be ignored when searching the database, but their values will be presented in the final result;
4. In the fields identified by the asterisk "\*", the user may select multiple choices to make the search more efficient;

5. After choosing the search criteria (Figure 2), click Run;
6. Next, the browser will be directed to a new page containing the results in table format (Figure 3);
7. Next, select the sequences of interest (Figure 4). If the user wants to select all sequences, the user should click in the first checkbox;
8. Next, the browser will be directed to the download page (Figure 5), where the sequence can be downloaded a FASTA or CSV file;
9. The downloaded sequences in FASTA format can be opened in almost all bioinformatics programs;
10. The CSV file can be imported into softwares (STATA, R) for statistical analysis of clinical and demographic information.

#### HTLV-1 subtyping tool

The HTLV-1 Subtyping Tool were developed using Java programming and PHP scripts. This tool accepts up to 1000 sequences at a time. In the first step of the analysis, the genomic region of reference sequence (ATK1 for HTLV-1) is identified using BLAST software. The second step involves the alignment of the query sequence with a complete reference dataset composed of all subtypes. The final step involves the construction of a phylogenetic tree using Tamura-Nei or HKY distance methods with a gamma distribution of among site rate heterogeneity, as implement in PAUP software [8].

A series of PHP scripts are used to read the XML output format produced by the JAVA program and create HTML report pages. The batch report contains information on the sequence name, length, assigned subtype and subgroup, and an illustration of the virus' genome. The HTLV-1 subtyping tool is hosted on the Africa Centre for Health and Population Studies bioinformatics group, UKZN, South Africa server with access at <http://www.bioafrica.net/regagenotype/html/indexhtlv.html/>.

#### Subtyping sequences

1. This tool uses phylogenetic methods to identify the subtype of query sequences.
2. Enter here your input data as FASTA format, after, click Run (Figure 6).

3. The batch report (Figure 7) will be the batch report and contain information on the sequence name, length, assigned subtype and a figure of the HTLV-1 genome. Accessing the report link will take the user to a report to each submitted sequence.

4. The sequence report (Figure 8) will be composed of three areas named: sequence assignment, analysis details and phylogenetic analyses.

5. The Sequence assignment contains information on (Figure 8):

The sequence submitted (name and length).

The classification assignment (subtype, subgroup and bootstrap support).

A graphical representation of the HTLV-1 genome showing the genomic region of the query sequence with the start and end positions related to the ATK1 genome.

The motivation of the classification (this is based on the decision tree).

6. The Phylogenetic analysis section contains (Figure 8):

The phylogenetic tree in PDF and Nexus format,

The log file generated by PAUP (contains info on the model of evolution and its parameters),

The alignment used.

#### Notes:

##### HTLV-1 Molecular Epidemiology Database

The access numbers in the output page are actual links to the respective sequence in the GenBank page.

The FASTA file features a header with the following structure:

>accession number, serial number, genomic region, isolate, base pair, subtype and subgroup (if available).

The CSV file is a table containing all information presented in the search screen, along with a column called Sequence. That column matches the CSV file serial number, making it possible to relate the information in the CSV file with that in the FASTA file.

## HTLV-1 subtyping tool

- The HTLV-1 subtyping is based only in the LTR region of the genome.
- The LASP HTLV-1 subtype tool accepts up to 1000 sequences at a time.

## References

1. Brucato N, Cassar O, Tonasso L, Tortevoeye P, Migot-Nabias F, Plancoulaine S, Guitard E, Larrouy G, Gessain A, Dugoujon J-M (2010) The imprint of the Slave Trade in an African American population: mitochondrial DNA, Y chromosome and HTLV-1 analysis in the Noir Marron of French Guiana. *BMC Evol Biol* 10:314
2. Van Prooyen N, Gold H, Andresen V, Schwartz O, Jones K, Ruscetti F, Lockett S, Gudla P, Venzon D, Franchini G (2010) Human T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission. *Proc Natl Acad Sci U S A* 107:20738–20743
3. Filippone C, Bassot S, Betsem E, Tortevoeye P, Guillotte M, Mercereau-Puijalon O, Plancoulaine S, Calattini S, Gessain A (2012) A new and frequent human T-cell leukemia virus indeterminate Western blot pattern: epidemiological determinants and PCR results in central African inhabitants. *J Clin Microbiol* 50:1663–72
4. Mota-Miranda ACA, Barreto FK, Amarante MFC, Batista E, Monteiro-Cunha JP, Farre L, Galvão-Castro B, Alcantara LCJ (2013) Molecular characterization of HTLV-1 gp46 glycoprotein from health carriers and HAM/TSP infected individuals. *Virology* 45:75
5. Araujo THA, Barreto FK, Luiz Carlos Júnior A, Miranda ACAM (2014) Inferences about the global scenario of human T-cell lymphotropic virus type 1 infection using data mining of viral sequences. *Mem Inst Oswaldo Cruz* 109:448–451
6. Araujo THA, Souza-Brito LI, Libin P, Deforche K, Edwards D, de Albuquerque-Junior AE, Vandamme A-M, Galvao-Castro B, Alcantara LCJ (2012) A Public HTLV-1 Molecular Epidemiology Database for Sequence Management and Data Mining. *PLoS One* 7:e42123
7. Alcantara LCJ, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, Galvão-Castro B, Vandamme AM, de Oliveira T (2009) A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res* 37:634–642
8. Wilgenbusch JC and Swofford D. (2003) Inferring evolutionary trees with PAUP. *Curr. Protoc. Bioinformatics*, Chapter 6, unit 6.4.

## Figures

Choose a criteria and make a search in our HTLV-1 Database.

The screenshot shows a search interface with the following elements:

- Genomic Region\***: A dropdown menu with options: env-pX, gag-pol-env-pX, pX, LTR, env.
- Sampling Date**: A dropdown menu.
- Subtype**: A dropdown menu.
- Continent\***: A dropdown menu with options: Africa, Asia, Central America, Europe, North America.
- Geographic Origin\***: A dropdown menu with options: Africa, Afro-caribbean, Algeria, Argentina, Bolivia.
- Gender**: A dropdown menu.
- Ethnicity**: A dropdown menu.
- Proviral Load**: A dropdown menu.
- CD4 Count**: A dropdown menu.
- CD8 Count**: A dropdown menu.
- Age**: A dropdown menu.
- Clinical Status**: A dropdown menu.

Below the filters, there is a note: **\*For multiple selections hold "ctrl"** and two buttons: **Clear** and **Run**.

Figure

1. HTLV-1 Molecular Epidemiology Database homepage displays interface.

Choose a criteria and make a search in our HTLV-1 Database.

This screenshot shows the same search interface as Figure 1, but with specific criteria selected and highlighted by black arrows:

- An arrow points to the **LTR** option in the **Genomic Region\*** dropdown.
- An arrow points to the **Asia** option in the **Continent\*** dropdown.
- An arrow points to the **Asymptomatic** option in the **Clinical Status** dropdown.

The **Clear** and **Run** buttons are visible at the bottom left.

Figure 2. Selection criteria to search in HTLV -1 Molecular Epidemiology Database homepage (example).

## HTLV-1 Database output

Information										
Your search has 7 results   Order by: <input type="text" value="Accession Number"/> <input type="button" value="Go"/>										
Please, select the sequences you need before clicking the download button. To select all the sequences, click in the first checkbox										
<input type="checkbox"/>	Accession Number	Genomic Region	Size (bp)	Gender	Age	Ethnicity	Geographic Origin	Continent	Clinical Status	CD4 Count (cells/ml)
<input type="checkbox"/>	S80215.1	LTR	292				Japan	Asia	Asymptomatic	
<input type="checkbox"/>	S80213.1	LTR	158				Japan	Asia	Asymptomatic	
<input type="checkbox"/>	S80212.1	LTR	184				Japan	Asia	Asymptomatic	
<input type="checkbox"/>	L42255.1	LTR	704	female	55		Kuwait	Asia	Asymptomatic	
<input type="checkbox"/>	AB211217.1	LTR	510				Iran	Asia	Asymptomatic	
<input type="checkbox"/>	AB211216.1	LTR	505				Iran	Asia	Asymptomatic	
<input type="checkbox"/>	AB211214.1	LTR	498				Iran	Asia	Asymptomatic	

Figure 3. Data presentation containing the search results.

<input type="checkbox"/>	Accession Number	Genomic Region	Size (bp)	Gender	Age	Ethnicity	Geographic Origin	Continent	Clinical Status
<input checked="" type="checkbox"/>	S80215.1	LTR	292				Japan	Asia	Asymptomatic
<input checked="" type="checkbox"/>	S80213.1	LTR	158				Japan	Asia	Asymptomatic
<input checked="" type="checkbox"/>	S80212.1	LTR	184				Japan	Asia	Asymptomatic
<input checked="" type="checkbox"/>	L42255.1	LTR	704	female	55		Kuwait	Asia	Asymptomatic

Figure 4. Selection of sequences of interest (Checkbox marking).

## HTLV-1 Download page

Information	
<a href="#">Download FASTA file</a>	<a href="#">Download CSV file</a>
<a href="#">Back to home</a>	<a href="#">Back to previous page</a>

Figure 5. Page to download the sequences in Fasta format and CSV file containing clinical, phylogenetics and epidemiological data.

### LASP HTLV-1 Automated Subtyping Tool (Version 1.0)

This tool uses phylogenetic methods to identify the subtype of query sequences.

*Please note:* The HTLV-1 subtyping is based only in the LTR region of the genome.

*Note for batch analysis:* The LASP HTLV-1 subtype tool accepts up to 1000 sequences at a time.

Enter here your input data as FASTA format.

[Choose a mirror to subtype your sequences](#) or [choose another virus to genotype](#).

```
>HTLV24_00085565
CGGGGGCTTAGAGCCTCCCAATGAAAAACATTTCCGCGAAACAGAGTCTBAAAAGG
TCA
BBGCCCCAACAATAAGCTCTGACCTCCCCCCCCBBAGGGACAGCTCAGCACCBBCTC
AGG
CTAGGCCCTGACGTGTCCCCCTGAAGACAAATGATAAGCTCAAGCTCCGGGAGGCC
```

Figure 6. HTLV-1 Subtyping Tool homepage interface: adding the sequence in FASTA format to run.



**HTLV Genotyping Tool Results**

Name	Length	Report	Assignment	Support	Genome
HTLV24_DQ005565_	719bp	<a href="#">Report</a>	subtype_a(subgroup_A)	99.0	
N3_Y16497_	719bp	<a href="#">Report</a>	subtype_a(subgroup_B)	99.0	
IDUSSA_DQ005555_	757bp	<a href="#">Report</a>	subtype_a(subgroup_A)	100.0	
BC11-2_U32552_	693bp	<a href="#">Report</a>	subtype_a(subgroup_A)	100.0	
test2	1190bp	<a href="#">Report</a>	subtype_a(subgroup_B)	100.0	
test	1318bp	<a href="#">Report</a>	HTLV1	NA	

Download results: [XML format](#), [CSV table](#)

Figure 7. The batch report that contain information on the sequence name, length, assigned subtype and a figure of the HTLV-1 genome. Accessing the report link will take the user to a report to each submitted sequence.

**Sequence Assignment**

**Sequence name : HTLV24\_DQ005565\_, length: 719 bps**  
**Assignment: subtype\_a(subgroup\_A), Bootstrap: 99.0%**

HTLV-1 Genome location ATK1      © LASP/FIOCRUZ HTLV1 Subtyping Tool

Your sequence start at position 53 and finish at position 769 in the ATK1 genome.  
 Motivation: *Subtype assigned based on sequence located in the LTR clustering with a HTLV1 subtype and/or subgroup with bootstrap > 60%*

Developed in cooperation with the [Evolutionary Biology Group](#) at University of Oxford, UK, the [REGA Institute](#) at the Katholieke Universiteit Leuven, Belgium and the [Laboratório Avançado de Saúde Pública \(LASP\)](#)/CPqGM/FIOCRUZ, Brazil.

**Analysis details**

- [Phylogenetic analyses](#)  
This section contains the alignments, inferred phylogenetic trees, and detailed results of the evolutionary analysis.

**Phylogenetic analyses**

**Phylogenetic analysis with pure subtypes:**

- Export or View the Phylogenetic Tree: [PDF](#), [NEXUS format](#).
- View the [PAUP\\* Log file](#) (Contains bootstrap values for all HIV subtypes)
- Download [the alignment \(Nexus format\)](#).

Figure 8. The sequence report will be composed of three areas named: sequence assignment, analysis details and phylogenetic analyses.

### 5.3 Artigo 2 - *manuscrito em preparação*: Caracterização molecular de novos genomas completos de HTLV-1 provenientes de pacientes com diferentes perfis clínicos

O HTLV-1 foi o primeiro retrovírus humano descrito e está associado a várias doenças como TSP/HAM, ATL e DIH. Estima-se que cerca de 5 a 10 milhões de pessoas estejam infectadas com HTLV-1 em todo o mundo, entretanto, um número inexpressivo de genomas completos estão disponíveis publicamente, cerca de 0,12 genomas completos por 10.000 indivíduos infectados e excepcionalmente em áreas consideradas endêmicas, como a cidade de Salvador, não existe genomas completos disponíveis. Além disso, há uma grande deficiência de informações clínicas e epidemiológicas em relação aos indivíduos dos quais as sequências foram oriundas. Assim, este estudo teve como objetivo gerar e caracterizar 31 sequências de genoma completos do HTLV-1 provenientes de indivíduos com TSP/HAM, ATL, DIH e assintomáticos, sendo a maioria dos indivíduos provenientes de Salvador.

Os resultados da caracterização molecular realizada nos genomas completos gerados e naqueles previamente disponíveis no *Genbank* sugerem que não há associação entre mutações e o perfil clínico. Detalhes deste estudo podem ser encontrados no manuscrito em preparação intitulado “***Molecular characterization of new HTLV-1 complete genomes from patients with different clinical forms***”.

#### **Molecular characterization of new HTLV-1 complete genomes from patients with different clinical forms**

Thessika Hialla Almeida Araújo<sup>1</sup>; Clayton Pereira Silva de Lima<sup>4</sup>; Rodrigo Santos de Oliveira<sup>4</sup>; Poliana da Silva Lemos<sup>4</sup>; Bernardo Galvão-Castro<sup>5</sup>; Simone Kashima Haddad<sup>6</sup>; Maria Lourdes Farre Vallve<sup>7</sup>; Achilea Candida Lisboa Bittencourt <sup>9</sup>; Edgar Marcelino de Carvalho Filho<sup>1,9</sup>; Aline Dórea Luz Menezes<sup>1</sup>; Fernanda Khouri Barreto<sup>1</sup>; Luciane Amorim Santos<sup>2</sup>; Filipe Ferreira de Almeida Rego<sup>3</sup>; Adjile Edjide Roukiyath Amoussa<sup>1</sup>; Marta Giovanetti<sup>1,10</sup>; Aline Cristina Andrade Mota Miranda Mascarenhas<sup>8</sup>; Márcio Roberto Teixeira Nunes<sup>4</sup>; Luiz Carlos Júnior Alcântara<sup>1,\*</sup>.

1 Centro de Pesquisa Gonçalo Moniz-Fiocruz, Laboratório de Hematologia Genética e Biologia Computacional, Salvador, Bahia, Brazil

2 Centro de Pesquisa Gonçalo Moniz-Fiocruz, Laboratório de Imunoparasitologia, Salvador, Bahia, Brazil

3 Universidade Católica do Salvador, Salvador, Bahia, Brazil

- 4 Instituto Evandro Chagas, Seção de Arbovirologia e Febres Hemorrágicas, Ananindeua, Pará, Brazil
- 5 Escola Bahiana de Medicina e Saúde Pública Salvador, Bahia, Brazil
- 6 Fundação Hemocentro de Ribeirão Preto, Laboratório de Biologia Molecular, Ribeirão Preto, São Paulo, Brazil
- 7 Centro de Pesquisa Gonçalo Moniz-Fiocruz, Laboratório da Patologia Experimental, Salvador, Bahia, Brazil
- 8 Universidade Federal da Bahia, Instituto de Ciências da Saúde, Departamento de Bioquímica e Biofísica, Salvador, Bahia, Brazil
- 9 Universidade Federal da Bahia, Brazil.
- 10 University of Rome "Tor Vergata"

\*Corresponding author

E-mail: [lalcan@bahia.fiocruz.br](mailto:lalcan@bahia.fiocruz.br)

## **Abstract**

The Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) is the first human retrovirus and is associated with several clinical syndromes as TSP/HAM, ATL and HID. It is estimated that approximately 5 to 10 million people are infected with HTLV-1, however the pathogenesis of some clinical manifestations is still not fully understood. Furthermore, there are few complete genomes publicly available, about 0.12 complete genomes per 10,000 infected individuals and the databases have a major deficiency of sequences information. Here, we generate and characterized 31 HTLV-1 complete genomes sequences derived from individuals with TSP/HAM, ATL, HID and asymptomatic patients, whose sequences have the information clinical and epidemiological about the patients. The 31 samples were amplified by PCR and the library was prepared and the sequencing reaction was performed by Ion Torrent. The sequencing data were assembled and mapped against the reference HTLV-1 genome. These sequences were submitted to the HTLV-1 Subtyping Tool, and all of them were genotyped as Cosmopolitan subtype, Transcontinental subgroup. The genetic distance, using the MEGA 7.0 software, showed low intragroup and intergroup diversity. The Geneious R6 software was used to analyze the non-coding region (LTR) to identify the transcriptions factors binding sites. Our analyses demonstrated that two LTR mutations were able to abolish the binding of Sp1 (transcription factor). The same software was used to identify the variants at genome coding regions. No statistical relationship between variants and different clinical profiles (TSP/HAM,

HID, ATL, Asymptomatic) were detected. Regarding the effect of mutations at the protein structure, the physico-chemical analyses and post-translational sites identification using the Network Protein Sequence Analysis (NPSA) and Prosite (GeneDoc) tools, respectively, were used. The physico-chemical analysis showed that one mutation at env and gag genes was able to decrease the antigenicity while other mutation at gag gene (p24) was able to increase the antigenicity. The post-translational modification analysis demonstrated a low frequency of mutations associated with the creation or abrogation of these sites in different clinical profiles, with no statistical association. In a second phase of the study, we searched the mutations found in the sequences generated in sequences previously published in Genbank, corroborating previous data, there was no statistical significance. This study contributed to increase of HTLV-1 complete genomes in world. Furthermore, to investigate better the contribution of HTLV-1 mutations for the disease outcome it is necessary evaluate the interaction of the viral genome and characteristics of the human host.

**Keywords:** HTLV-1; complete genome; next generation sequencing.

## Background

Nearly four decades ago, the Human T-Cell Lymphotropic Virus Type 1 (HTLV-1) was isolated and described [1]. The HTLV-1 was the first human retrovirus and is associated with several diseases: Tropical Spastic Paraparesis/HTLV-1-Associated Myelopathy (TSP/HAM) Adult T-cell leukemia/lymphoma (ATL) and other inflammatory diseases, such as HTLV-1 associated infective dermatitis (HID) [2–4].

It is estimated that approximately 5 to 10 million people are infected with HTLV worldwide. This distribution is heterogeneous, especially the endemic regions, such as Southwest Japan, the Caribbean Islands, South America and Central Africa [5]. Studies have shown that Brazil is considerate an endemic area for HTLV, where the virus prevalence varies depending on the geographic region and the population analyzed [6–8]. According to a population-based study, Salvador, Bahia is an important area with a prevalence of 1.76% (reaching 8.4% in women aged 51 years or above). Besides the higher prevalence in women, the virus was associated with lower education and lower income levels, as well as areas with the worst indicators of socioeconomic status [9].

The HTLV-1 genome is composed by the genes gag, pol, env and pX region flanked by two long terminal repeat (LTR) regions at both 5' and 3' ends, which contains enhancer/promoter genetic elements, that are critical in viral RNA transcription. The LTR has a high genetic variability and is used in molecular epidemiology studies to classify the HTLV-1 subtypes. Seven HTLV-1 subtypes was described (a-g), while the Cosmopolitan subtype (a) is worldwide distributed [10–19]. Cosmopolitan subtype has experimented a degree of molecular diversity giving rise to five subgroups: Transcontinental (A), Japanese (B), West African (C), North African (D), Black Peruvian (E) and Ethiopia (F) [13,20–23].

Despite years of study, the pathogenesis of some clinical manifestations is still not fully understood. Furthermore, there are few complete genomes publicly available (0.12 complete genomes per 10,000 infected individuals), exceptionally in areas considered endemic, such as the city of Salvador (Bahia/Brazil), there are no complete genomes available.

Furthermore, the databases have a major deficiency of sequences information, including

data of proviral load values, clinical profile, sampling date and other important data. Here, we generate 31 HTLV-1 complete genomes sequences derived from individuals with different clinical profile, whose sequences have all information about the patients. We performed molecular analysis to characterize and compare our sequences with sequences already published.

## **Material and Methods**

### **Patients**

The study included 31 samples from HTLV-1 patients with defined clinical profiles: 10 samples from patients with TSP/HAM, 7 samples from ATL patients, 9 samples from asymptomatic patients and 5 samples from patients with HID. These patients were from Center of HTLV-1 Bahia School of Medicine and Public Health – Salvador, Bahia, Hospital Complex Prof. Edgar Santos – Salvador, Bahia and from Hemocentro - Ribeirão Preto. All samples were anonymized and informed consent was written and obtained from each subject. The clinical classification was carried out by medical experts according to World Health Organization (WHO) and the research was approved by the Ethics Committee of the Centro de Pesquisa Gonçalo Moniz/ Fiocruz.

### **Amplification of proviral genome**

To amplify HTLV-1 products to generate the complete genomes by Ion Torrent (Life Technologies), first we used all HTLV-1 genome complete available in Genbank to create a consensus sequence. Through the online package IDT SciTools® Web Tools we generated 32 pairs of primers capable of amplifying the most prevalent variant of HTLV-1 (Supplementary table 1). Each pair of primers is capable of amplifying a fragment of 400 bases pairs with an overlap 50 bases pairs. The PCR conditions were as follows: denaturation at 94°C for 3 minutes, annealing at 94°C (15 sec), 65 °C (45 sec), 72 °C (1 minute) 35 times of cycle, a final extension of 72°C for 8 minutes. The amplified products

were analyzed on 1% agarose gel and quantified using Qubit™ dsDNA HS Assay Kit (Invitrogen).

### **Library preparation and sequencing**

The amplicons of each patient were pooled and later were fragmented using the Bioruptor® standard (Diagenode). The libraries were built by the automated AB Library Builder System (Applied Biosystems). To normalize the number of molecules required for the emulsion PCR, a quantitation step was performed using the automated Ion OneTouch 2 platform. Ion Xpress™ Adapters and unique Ion Xpress™ Barcodes (Life Technologies) were ligated to fragmented material by using the Ion Plus Fragment Library kit, followed by the Emulsion PCR (ePCR) and enrichment steps that were performed using Ion PGM OT2 200 Template Kit (Thermo Fisher Scientific) according to the manufacturer's protocol (Ion 2 platform OT2 OneTouch System™ - Life Technologies). Finally, the mix was loaded on a 318v2 chip and the sequencing reaction was performed on the PGM device.

### **Data analysis**

To perform the molecular analysis, the fastaq files from each viral genome generated was trimmed and mapped to the HTLV-1 reference sequence ATK-1 (J02029), using the highest sensitivity and the fine tuning, followed by edition and manual alignment of the mapped reads.

First, the genetic distances among the sequences were measured using the MEGA 7.0 program within and between the four distinct groups: asymptomatic, ATL, HID, and TSP/HAM. Then, the LTR sequences were submitted to the HTLV-1 Automated Subtyping Tool Version 1.0 to identify the HTLV-1 subtype of the genomes [24].

To phylogenetic analysis, the new genomes and one reference from each HTLV-1 subtype sequences were aligned using [25]. The HKY85 model of nucleotide substitution was used, as suggested by jModelTest, and the maximum likelihood tree was inferred using PhyML tool. The bootstrap analysis (1000 replicates) was used to calculate the statistical support of the tree branches. Tree visualization was done using Geneious R6 software.

The non-coding region (LTR) was used to search motifs for transcription factors using TFScan plugin of the Geneious R6 software. First, the major motifs were annotated: Tax responsible elements - TxRE (three 21-pair base imperfect repetitions), TATA BOX (TATAA) and E-Box (CANNTG). Then, we search for nucleotide variants that possibly influence the binding of transcription factors at these motifs.

The coding regions analysis was done through the search of variants using the Geneious R6 Software. The statistical analyses were conducted using the STATA program, with Fisher nonparametric test and a P value lower than 0.05 was considered statistically significant. We also evaluated the possible influence of these variants on the protein structure through the amino acid family change analysis, physico-chemical analysis and search for post-translational modifications sites [26]. The physico-chemical analysis was performed using Network Protein Sequence Analysis (NPSA) (<http://npsa-pbil.ibcp.fr/>) and the search for post-translational modifications sites was done using the GeneDoc software and the Prosite tool. Next, we search the information finding in the 31 complete genomes generated in all HTLV-1 Cosmopolitan Transcontinental complete and partial sequences with clinical forms data available on GenBank database (accession numbers): AF042071.1, AF259264.1, AB979451.1, M86840.1, L36905.1, KF797815.1 - KF797835.1, KF797839.1 - KF797850.1, KF797852.1, KF797853.1, KF797855.1, KF797857.1, KF797860.1 - KF797873.1, KF797875.1 - KF797882.1, KF797884.1 - KF797886.1, KF797888.1, KF797889.1, DQ005564.1 - DQ005566.1, AY342300.1 - AY342302.1, AY342305.1, AY342306.1, DQ512875.1, DQ512874.1, DQ471187.1 - DQ471209.1, AY499185.1, AB036351.1, AB036350.1, AB036348.1, L42255.1, AB211217.1, AB211216.1, AB211214.1, AF124043.1, AF014664.1 - AF014670.1, AF014662.1, AF014659.1, AF014653.1 - AF014657.1, GU126475.1, GU126478.1, EU622583.1, EU622582.1, EU392164.1, DQ512873.1, DQ897682.1 - DQ897686.1, AB036349.1, AF014644.1 - AF014646.1, U25140.1, U25139.1, U25062.1, L42225.1, AB211220.1, AB211218.1, AB211215.1, AF097525.1 - AF097528.1, AF014647.1 - AF014652.1, GU126474.1, GU126476.1, GU126479.1 - GU126483.1, L42253.1.1, GU126487.1, DQ663637.1, GU126484.1, GU126488.1, GU126489.1,



DQ663642.1, GU126485.1, X56949.2, AB036377.1.1, AB211204.1, AB211201.1 and AF485381.1.

### **Datamining**

We used the Bayesian network (BN) learning (using B-course software adapted by Deforche et al. (2006)) was used to describe and visualize conditional dependencies between the variables [27]. Dependencies are represented qualitatively by a directed acyclic graph where each node corresponds to a variable, and a direct arc between nodes represents a direct influence. Robustness of the arcs was scored with a nonparametric bootstrap test (100x replicates) and only arcs with more than 30% support were depicted. Included variables were clinical profiles (Asymptomatic, ATL, TSP/ HAM and HID) and different mutations found in the noncoding regions of the genome [28].

### **Supporting data**

All sequences newly reported here has been deposited in the NCBI's GenBank database (accession numbers KY007244-KY007274).

### **Results**

The study population consisted of 31 patients: HID (n=5); TSP/HAM (n=10); asymptomatic (n=9) and ATL (n=7). The age of patients ranged between 12 and 67 years, with median of 41 years. Females represented 61% of the samples. The median proviral loads from patients was measured: TSP/HAM (117 copies per  $10^3$  PBMCs); HID (99 copies per  $10^3$  PBMCs); ATL (419 copies per  $10^3$  PBMCs) and asymptomatic (41 copies per  $10^3$  PBMCs).

### **Molecular characterization**

We generated 31 HTLV-1 complete genome sequences with 8998 bases pairs. These sequences were classified as subtype a (cosmopolitan), subgroup A (transcontinental) (Figure 1). The genetic distances were measured, within and between the four groups with different clinical profiles (TSP/HAM, asymptomatic, HID and ATL). These results are shown in Table 1.

### **Non-coding region analysis**

Considering that the transcription factors motifs are important to regulate the transcription of HTLV-1, mutations associated to the loss or gain of any of motifs could be responsible for changes in the viral gene expression. Here, we investigated mutations at the TxRE, TATA BOX and E-Box from HTLV-1 LTR region and observed only 2 mutations at the TxRE that may abrogate Sph1 binding site. The A125G mutation, found in 4 TSP/HAM, 3 ATL and 4 asymptomatic sequences may abrogate the TxRE1, while G174A mutation, found in 4 TSP/HAM, 2 ATL and 3 asymptomatic sequences may abrogate the TxRE2.

Then we search these two mutations in all LTR Cosmopolitan Transcontinental sequences with clinical forms data available on GenBank database. The A125G mutation was also found in 15 TSP/HAM, 26 asymptomatic and 1 ATL sequences while G174A mutation was found only in 17 TSP/HAM, 31 asymptomatic and 1 ATL sequences. Similar to our findings, the A125G and G174A mutations were not found in sequences with the clinical profile HID.

### **Coding region analysis**

We identified 377 variants at the coding regions, whose 226 were synonymous (no amino acid change). 151 non-synonymous mutation (change the amino acid) were identified in gag (n=28), pro (n=9), pol (n=26), env (n=37) and pX region (n=51). To evaluate the possible influence of these mutations on HTLV-1 proteins we analyzed if these non-synonymous mutations result in amino acid family change and found a total of 63 mutations that could alter the amino acid family. These 63 mutations were distributed in the entire genome and we analyzed the frequency of these mutations in each coding region regarding to clinical profiles (Figure 2A). To reveal possible interactions among mutations and clinical profiles we submitted to datamining analyses the non-synonymous mutations that could alter the amino acid family found with frequency of at least 5%. No association was found between the 12 evaluated mutations and clinical profile (Figure 2B).

After determining the variation, we started the characterization of post-translational modifications sites and physico-chemical profiles (Table 2). We identified 6 mutations in the analyzed proteins that were able to create post-translational modifications sites while 4 mutations were able to abolish post-translational modifications sites. For the physico-chemical analysis we selected mutations found in more than 5% of the samples and mutations described in p13, p15, pro, pol, gp21, p12, p30, tax and rex that did not cause changes when compared to wild-type sequence. However, the mutations V247I (gp46) and T106I (p19) were able to reduce antigenicity. Only I153V mutation found in p24 increased antigenicity.

We have expanded our analysis searching in all HTLV-1 Cosmopolitan Transcontinental sequences with clinical forms data available on GenBank database the 10 mutations that may create or abolish some of post-translational modifications and the 3 mutations that may alter the physico-chemical (Supplementary table 2).

## **Discussion**

The next generation sequencing (NGS) are high throughput sequencing technologies capable of generating information about millions base pairs in a single run. These new platforms generate a greater number of data, in a shorter period of time and with a lower cost per based sequenced [29,30]. The researchers with expertise in HTLV-1 raised the need for further studies involving genetic characterization of complete genomes that could help understand virus pathogenicity and the development of disease [11,31–44]. Here, we generated and characterized 31 HTLV-1 complete genomes from asymptomatic, ATL, TSP/HAM and HID patients. The sequences from patients with HID are the first worldwide complete genomes publicly available.

As previous studies, the phylogenetic analysis showed the cosmopolitan subtype transcontinental subgroup (aA) are the most prevalent genotype circulating in Brazil [8,9]. The genetic distances between intragroup and intergroup demonstrated low diversity and the mutations frequency found in the HTLV-1 complete genomes generated here were not

statistically significant between different clinical profiles as in analysis of others available sequences of GenBank.

We also analyzed each mutation in the non-coding region and in the coding region. Two LTR mutations were able to abolish Sp1 binding site in specific TxREs. The Sp1 factor is critical for basic leucine zipper factor (HBZ) transcription, which plays a significant role in the proliferation of infected cells [45]. Furthermore, we identified two mutations in the coding region that were able to reduce antigenicity and one that increase antigenicity. We cannot associate these results found with the clinical profile, due to the reduced number of individuals recruited in this study. However, these mutations may to become biomarkers as they are located in important protein domains.

The post-translational modification events play an important role in viral replication and cell transformation. Our analyses demonstrated a high identity among the sequences, suggesting the influence of these sites on the host immune response, and virus persistence. We identified a low frequency of mutations associated with the creation or abrogation of these sites in different clinical profiles. Some modifications as phosphorylation, ubiquitination and acetylation are critical for Tax transactivation via both ATF/CREB and NF- $\kappa$ B pathways and this could act in inhibition of DNA repair, cell cycle control and activation of p53 tumor suppressor [46–54]. The post-translational modifications in other proteins are not well documented, therefore there is a lack of evidence of the importance of these sites for viral infection [55].

## **Conclusions**

This study contributed to increase of HTLV-1 complete genomes published with information about clinical profile and the results of molecular characterization suggest no association between mutations and clinical profile. To investigate better, the contribution of HTLV-1 mutations for the disease outcome it is necessary evaluate the interaction of the viral genome and characteristics of the human host. Furthermore, this study contributed to increase of HTLV-1 complete genomes published and this may help researches that search understand virus pathogenicity and development of HTLV-1 disease.

## Acknowledgments:

We would like to thank the FAPESB financial support.

## References

1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* 1980;77:7415–9.
2. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet Lond. Engl.* 1985;2:407–10.
3. La Grenade L, Manns A, Fletcher V, Derm D, Carberry C, Hanchard B, et al. Clinical, pathologic, and immunologic features of human T-lymphotropic virus type I-associated infective dermatitis in children. *Arch. Dermatol.* 1998;134:439–44.
4. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc. Natl. Acad. Sci. U. S. A.* 1982;79:2031–5.
5. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front. Microbiol.* 2012;3:388.
6. Catalan-Soares B, Carneiro-Proietti AB de F, Proietti FA, Interdisciplinary HTLV Research Group. Heterogeneous geographic distribution of human T-cell lymphotropic viruses I and II (HTLV-I/II): serological screening prevalence rates in blood donors from large urban areas in Brazil. *Cad. Saúde Pública.* 2005;21:926–31.
7. Galvão-Castro B, Loures L, Rodrigues LG, Sereno A, Ferreira Júnior OC, Franco LG, et al. Distribution of human T-lymphotropic virus type I among blood donors: a nationwide Brazilian study. *Transfusion (Paris).* 1997;37:242–3.
8. Carneiro-Proietti ABF, Sabino EC, Leão S, Salles NA, Loureiro P, Sarr M, et al., NHLBI Retrovirus Epidemiology Donor Study-II (Reds-II), International Component. Human T-lymphotropic virus type 1 and type 2 seroprevalence, incidence, and residual transfusion risk among blood donors in Brazil during 2007-2009. *AIDS Res. Hum. Retroviruses.* 2012;28:1265–72.
9. Dourado I, Alcantara LCJ, Barreto ML, da Gloria Teixeira M, Galvão-Castro B. HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. *J. Acquir. Immune Defic. Syndr.* 1999. 2003;34:527–31.
10. Hahn BH, Shaw GM, Popovic M, Lo Monico A, Gallo RC, Wong-Staal F. Molecular cloning and analysis of a new variant of human T-cell leukemia virus (HTLV-ib) from an African patient with adult T-cell leukemia-lymphoma. *Int. J. Cancer.* 1984;34:613–8.
11. Gessain A, Boeri E, Yanagihara R, Gallo RC, Franchini G. Complete nucleotide sequence of a highly divergent human T-cell leukemia (lymphotropic) virus type I (HTLV-I) variant from melanesia: genetic and phylogenetic relationship to HTLV-I strains from other geographical regions. *J. Virol.* 1993;67:1015–23.

12. Bastian I, Gardner J, Webb D, Gardner I. Isolation of a human T-lymphotropic virus type I strain from Australian aboriginals. *J. Virol.* 1993;67:843–51.
13. Miura T, Fukunaga T, Igarashi T, Yamashita M, Ido E, Funahashi S, et al. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc. Natl. Acad. Sci. U. S. A.* 1994;91:1124–7.
14. Miura T, Yamashita M, Zaninovic V, Cartier L, Takehisa J, Igarashi T, et al. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. *J. Mol. Evol.* 1997;44 Suppl 1:S76–82.
15. Vandamme AM, Liu HF, Goubau P, Desmyter J. Primate T-lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-I. *Virology.* 1994;202:212–23.
16. Chen J, Zekeng L, Yamashita M, Takehisa J, Miura T, Ido E, et al. HTLV type I isolated from a Pygmy in Cameroon is related to but distinct from the known central African type. *AIDS Res. Hum. Retroviruses.* 1995;11:1529–31.
17. Mahieux R, Ibrahim F, Mauclere P, Herve V, Michel P, Tekaiia F, et al. Molecular epidemiology of 58 new African human T-cell leukemia virus type 1 (HTLV-1) strains: identification of a new and distinct HTLV-1 molecular subtype in Central Africa and in Pygmies. *J. Virol.* 1997;71:1317–33.
18. Salemi M, Van Dooren S, Audenaert E, Delaporte E, Goubau P, Desmyter J, et al. Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have closest relatives among African STLV-I strains. *Virology.* 1998;246:277–87.
19. Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, et al. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102:7994–9.
20. Gasmi M, Farouqi B, d' Incan M, Desgranges C. Long terminal repeat sequence analysis of HTLV type I molecular variants identified in four north African patients. *AIDS Res. Hum. Retroviruses.* 1994;10:1313–5.
21. Vidal AU, Gessain A, Yoshida M, Tekaiia F, Garin B, Guillemain B, et al. Phylogenetic classification of human T cell leukaemia/lymphoma virus type I genotypes in five major molecular and geographical subtypes. *J. Gen. Virol.* 1994;75 ( Pt 12):3655–66.
22. Van Dooren S, Gotuzzo E, Salemi M, Watts D, Audenaert E, Duwe S, et al. Evidence for a post-Columbian introduction of human T-cell lymphotropic virus [type I] [corrected] in Latin America. *J. Gen. Virol.* 1998;79 ( Pt 11):2695–708.
23. Treviño A, Alcantara LC, Benito R, Caballero E, Aguilera A, Ramos JM, et al., HTLV Spanish Study Group. Molecular epidemiology and clinical features of human T cell lymphotropic virus type 1 infection in Spain. *AIDS Res. Hum. Retroviruses.* 2014;30:856–62.
24. Alcantara LCJ, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res.* 2009;37:W634–42.
25. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.

26. Exarchos KP, Exarchos TP, Papaloukas C, Troganis AN, Fotiadis DI. Detection of discriminative sequence patterns in the neighborhood of proline cis peptide bonds and their functional annotation. *BMC Bioinformatics*. 2009;10:113.
27. Deforche K, Silander T, Camacho R, Grossman Z, Soares MA, Van Laethem K, et al., non-B Workgroup. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinforma. Oxf. Engl.* 2006;22:2975–9.
28. Friedman N, Goldszmidt M, Wyner A. Data Analysis with Bayesian Networks: A Bootstrap Approach. *ArXiv13016695 Cs Stat [Internet]*. 2013 [cited 2016 Apr 4]; Available from: <http://arxiv.org/abs/1301.6695>
29. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* 2013;19:15–22.
30. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* 2012;93:1853–68.
31. Bazarbachi A, Huang M, Gessain A, Saal F, Saib A, Peries J, et al. Human T-cell-leukemia virus type I in post-transfusional spastic paraparesis: complete proviral sequence from uncultured blood cells. *Int. J. Cancer.* 1995;63:494–9.
32. Andonov A, Coulthart MB, Pérez-Losada M, Crandall KA, Posada D, Padmore R, et al. Insights into origins of Human T-cell Lymphotropic Virus Type 1 based on new strains from aboriginal people of Canada. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 2012;12:1822–30.
33. Malik KT, Even J, Karpas A. Molecular cloning and complete nucleotide sequence of an adult T cell leukaemia virus/human T cell leukaemia virus type I (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/HTLV-I subgroup. *J. Gen. Virol.* 1988;69 ( Pt 7):1695–710.
34. Zhao TM, Robinson MA, Sawadikosol S, Simpson RM, Kindt TJ. Variation in HTLV-I sequences from rabbit cell lines with diverse in vivo effects. *Virology.* 1993;195:271–4.
35. Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl. Acad. Sci. U. S. A.* 1983;80:3618–22.
36. Ellerbrok H, Fleischer C, Salemi M, Reinhardt P, Ludwig WD, Vandamme AM, et al. Sequence analysis of the first HTLV-I infection in Germany without relations to endemic areas. *AIDS Res. Hum. Retroviruses.* 1998;14:1199–203.
37. Chou KS, Okayama A, Tachibana N, Lee TH, Essex M. Nucleotide sequence analysis of a full-length human T-cell leukemia virus type I from adult T-cell leukemia cells: a prematurely terminated PX open reading frame II. *Int. J. Cancer.* 1995;60:701–6.
38. Cao F, Ji Y, Huang R, Zhao T, Kindt TJ. Nucleotide sequence analyses of partial envgp46 gene of human T-lymphotropic virus type I from inhabitants of Fujian Province in Southeast China. *AIDS Res. Hum. Retroviruses.* 2000;16:921–3.
39. Ohsugi T, Kumasaka T, Urano T. Construction of a full-length human T cell leukemia virus type I genome from MT-2 cells containing multiple defective proviruses using overlapping polymerase chain reaction. *Anal. Biochem.* 2004;329:281–8.

40. Zanella L, Otsuki K, Marin MA, Bendet I, Vicente AC. Complete genome sequence of Central Africa human T-cell lymphotropic virus subtype 1b. *J. Virol.* 2012;86:12451.
41. Cassar O, Einsiedel L, Afonso PV, Gessain A. Human T-cell lymphotropic virus type 1 subtype C molecular variants among indigenous australians: new insights into the molecular epidemiology of HTLV-1 in Australo-Melanesia. *PLoS Negl. Trop. Dis.* 2013;7:e2418.
42. Kuramitsu M, Okuma K, Yamagishi M, Yamochi T, Firouzi S, Momose H, et al. Identification of TL-Om1, an adult T-cell leukemia (ATL) cell line, as reference material for quantitative PCR for human T-lymphotropic virus 1. *J. Clin. Microbiol.* 2015;53:587–96.
43. Martin F, Bangham CRM, Ciminale V, Lairmore MD, Murphy EL, Switzer WM, et al. Conference highlights of the 15th International Conference on Human Retrovirology: HTLV and related retroviruses, 4-8 June 2011, Leuven, Gembloux, Belgium. *Retrovirology.* 2011;8:86.
44. Ma Y, Zheng S, Wang N, Duan Y, Sun X, Jin J, et al. Epidemiological analysis of HTLV-1 and HTLV-2 infection among different population in Central China. *PloS One.* 2013;8:e66795.
45. Yoshida M, Satou Y, Yasunaga J-I, Fujisawa J-I, Matsuoka M. Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene. *J. Virol.* 2008;82:9359–68.
46. Lodewick J, Lamsoul I, Polania A, Lebrun S, Burny A, Ratner L, et al. Acetylation of the human T-cell leukemia virus type 1 Tax oncoprotein by p300 promotes activation of the NF-kappaB pathway. *Virology.* 2009;386:68–78.
47. Lodewick J, Lamsoul I, Bex F. Move or die: the fate of the Tax oncoprotein of HTLV-1. *Viruses.* 2011;3:829–57.
48. Durkin SS, Ward MD, Fryrear KA, Semmes OJ. Site-specific phosphorylation differentiates active from inactive forms of the human T-cell leukemia virus type 1 Tax oncoprotein. *J. Biol. Chem.* 2006;281:31705–12.
49. Peloponese J-M, Iha H, Yedavalli VRK, Miyazato A, Li Y, Haller K, et al. Ubiquitination of human T-cell leukemia virus type 1 tax modulates its activity. *J. Virol.* 2004;78:11686–95.
50. Lamsoul I, Lodewick J, Lebrun S, Brasseur R, Burny A, Gaynor RB, et al. Exclusive ubiquitination and sumoylation on overlapping lysine residues mediate NF-kappaB activation by the human T-cell leukemia virus tax oncoprotein. *Mol. Cell. Biol.* 2005;25:10391–406.
51. Nasr R, Chiari E, El-Sabban M, Mahieux R, Kfoury Y, Abdulhay M, et al. Tax ubiquitylation and sumoylation control critical cytoplasmic and nuclear steps of NF-kappaB activation. *Blood.* 2006;107:4021–9.
52. Kfoury Y, Nasr R, Favre-Bonvin A, El-Sabban M, Renault N, Giron M-L, et al. Ubiquitylated Tax targets and binds the IKK signalosome at the centrosome. *Oncogene.* 2008;27:1665–76.
53. Kfoury Y, Nasr R, Journo C, Mahieux R, Pique C, Bazarbachi A. The multifaceted oncoprotein Tax: subcellular localization, posttranslational modifications, and NF- $\kappa$ B activation. *Adv. Cancer Res.* 2012;113:85–120.
54. Jeang KT. Functional activities of the human T-cell leukemia virus type I Tax oncoprotein: cellular signaling through NF-kappa B. *Cytokine Growth Factor Rev.* 2001;12:207–17.



55. Kannagi M, Shida H, Igarashi H, Kuruma K, Murai H, Aono Y, et al. Target epitope in the Tax protein of human T-cell leukemia virus type I recognized by class I major histocompatibility complex-restricted cytotoxic T cells. *J. Virol.* 1992;66:2928–33.

## Figure

Figure 1: Maximum Likelihood tree based on HTLV-1 entire genome sequences. Sequences of the cosmopolitan subtype and transcontinental subgroup are in red. Numbers besides internal branches indicate bootstrap values based on 1000 replicates.

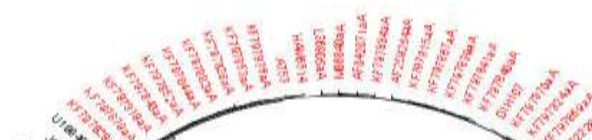
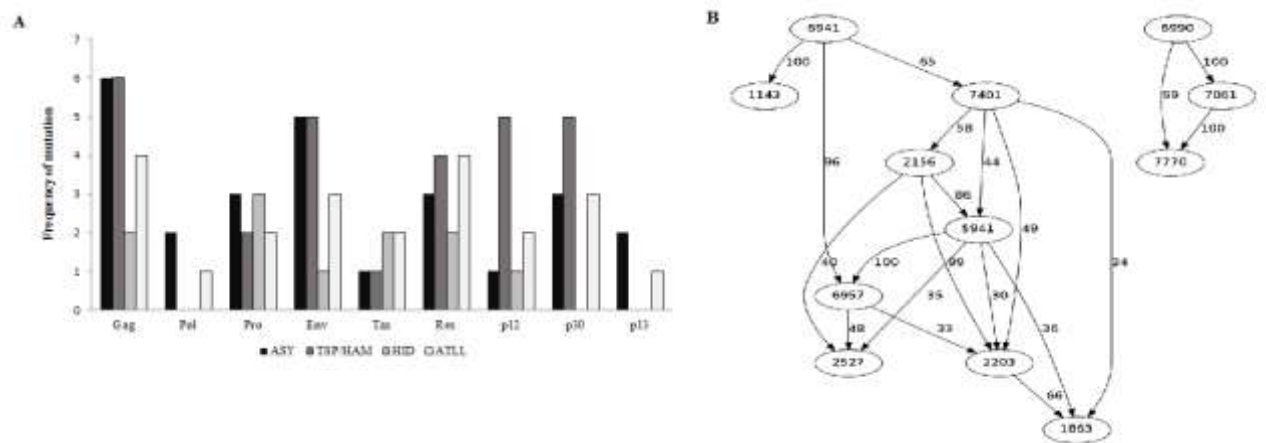


Figure 2: Analysis of non-synonymous mutations that alter amino acid families found in HTLV-1 complete genome.



(A) Distribution of 63 mutations according to the clinical status and genomic region. The Fisher nonparametric test was conducted and no statistical significance was observed. (B) Analysis of interactions among 12 mutations found with frequency of at least 5% in HTLV-1 complete genome.

## Tables

Table 1: Genetic distances in different clinical forms of HTLV-1 complete genomes using the MEGA 7.0 program.

	HID	TSP/HAM	ATL	Asymptomatic
HID	0.006	0.007	0.006	0.006
TSP/HAM	0.007	0.008	0.007	0.007
ATL	0.006	0.007	0.007	0.006
Asymptomatic	0.006	0.007	0.006	0.006

Table 2: Physico-chemical and post translational modifications analysis of 31 HTLV-1 complete genome.

Gene (protein)	Physico-chemical analysis	Post translational modifications	Clinical Profile Frequency (%)				p-value
			ASY n=9	HID n=5	TSP/HAM n=10	ATL n=7	
<b>Gag (p15)</b>							
P8S	-	Creation cAMP- and cGMP-dependent protein kinase phosphorylation site	1	1	-	-	0.389
P77R	-	Creation cAMP- and cGMP-dependent protein kinase phosphorylation site	-	1	-	-	1.0
<b>Gag (p19)</b>							
T106I	Reduced antigenicity	-	-	-	1	3	0.052
<b>Gag (p24)</b>							
I153V	Increased antigenicity	-	1	1	-	-	0.146
<b>Pro</b>							
G201D	-	Abrogation N-myristoylation site	-	-	1	-	0.538
<b>Pol</b>							
T40A	-	Abrogation protein kinase C phosphorylation site	-	-	-	1	0.315
Q446R	-	Creation protein kinase C phosphorylation site	3	1	1	1	0.614
<b>Env (gp46)</b>							
W88S	-	Creation protein kinase C phosphorylation site	-	-	-	1	0.315
P92S	-	Creation protein kinase C phosphorylation site	-	-	1	-	0.538
V247I	Reduced antigenicity	-	5	-	4	2	0.206
G307R	-	Creation protein kinase C phosphorylation site	-	1	-	-	0.146
<b>pX (Tax)</b>							
D264G	-	Creation N-myristoylation site	-	-	1	-	0.538
<b>pX (p13)</b>							
R47C	-	Abrogation protein kinase C phosphorylation site	1	-	-	-	0.471
<b>pX (p30)</b>							
R37C	-	Abrogation protein kinase C phosphorylation site	-	1	5	3	0.083

Position related to J02029. Asy, asymptomatic; HID, HTLV-1 associated infective dermatitis; TSP/HAM, Tropical Spastic Paraparesis /HTLV-1-Associated Myelopathy; ATL, Adult T-cell leukemia/lymphoma. P- values were calculated by the Fisher's exact test.

## Supplementary table

Supplementary table 1: Primers sequences used in the study to amplify the HTLV-1 complete

genome. \*F – Forward \*\* R – Reverse

N° of Fragment	Primers sequence
1	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GGC TTA GAG CCT CCC AGT G -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAG AAC GCG ACT CAA CCG -3'
2	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GGA AGA CAA ATC ATA AGC TCA GAC CT -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC GGC AGT TAG TCG TGA AT -3'
3	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAG AGC TCA GGT CGA GAC C -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC GGT GTT TCT AAA GCT ATT TTA AGA A -3'
4	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAC AGC CCA TTC TAT AGC ACT CT -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTG CAT GAC TGG AAG GAC TTG -3'
5	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAC TCA TCC AAA CCC AAG CC -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTG GTA TTC TCG CCT TAA TCC TTG -3'
6	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAA AGA CCT CCA AGA CCT CCT -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGG ACA ACT AAC ACT TTG GTT TTG T – 3'
7	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCT TCG TAG AAC GCC TCA ACA TA -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAA GGA CTT GCT GTA ATG TGG G -3'
8	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GTC AAG ACC CAA CTC ACT GGA -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTT TTG GTA TCA ACT AGG CAA GAT GT -3'
9	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAA AGC CCA GGT TGA CAC C -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC GGT GTA GGG TTC GAT ATG -3'
10	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC GCC TGT TAT CTT GCC AA -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAT TTC GAA CAG GGT GGG AC -3'
11	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCA GGA AAT AAC CCA GTA TTC CCA -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGT TTT CGG ACA CAG GCA AC -3'
12	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAA ATG CAG CTG GCC CAT A -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAG TCT ACT GCG GCA GTT CT -3'
13	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAG CCC CTT CAC AGT CTC T -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAA AGT GTT CCA AAG TTC TCC AGT -3'
14	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAC TTG CCT CAG CTG TGT TAT TAC -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC GAA GGT AAT GAT AAA GAT ACT TGG A -3'
15	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GTT TCC CCG GTG ATC ATA AAC AC -3' R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTT GCC CCT TGC AAT GTG A -3'
16	F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC CGC TTA CTA TCG CGT AA -3'

**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC GTT GTC TGT GTT TAT GTA GC -3'**  
 17 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAC CCA CAA CAT CAG ATG CC -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGT TTT GTG GCA GTT GGT TAA CA -3'**  
 18 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCT CCC TTG CTA TTC GCC A -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAA ACT TAC CCA TGG TGT TGG T -3'**  
 19 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAA CTG CCA CAA AAC CCG A -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTG TAG GGC CTG ATC TGC T -3'**  
 20 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAC CAA CAC CAT GGG TAA GTT T -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAG GGA AAA CCG CAT TTC GAA -3'**  
 21 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GTC AGC CTC TTA TTC AGA CCC TT -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTG GAA GCG CTA ACG ATG G -3'**  
 22 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GTC GAT CGT GCC AGC CTA T -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCT TGC TCC CAG AAC AGG A -3'**  
 23 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC CTG TCA TAA CTC CCT CAT C -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAC AAG GGT GAT TCC AGT TTG TAA -3'**  
 24 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAG TAT GCT GCC CAG AAC AG -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAG GTG GAG TCC TTG GAG G -3'**  
 25 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GAT TGC AAC CAC ATC GCC T -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCT TGA CGG TTT GCT ATC CTT AGA -3'**  
 26 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC TGC TCT TCC TGC TTT CT -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GTT GTA CAC AGT CTC CAA ACA CG -3'**  
 27 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCT TGT CCT TTA ACT CTT CCT CCA -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGG TTC CAT GTA TCC ATT TCG GA -3'**  
 28 **F 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGC AAA CAG TCT TCG GGT AGA ATT AT -3'**  
**R 5'- CGT ATC GCC TCC CTC GCG CCA TCA GTC CAG AGC ATC AGA TCA CCT -3'**  
 29 **F 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAC TGT ATG AGG CCG TGT GA -3'**  
**R 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC CAA AAC CTG TAC ACC CT -3'**  
 30 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCC AGC CAT CTT TAG TAC TAC AGT -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GCC ATA TGC GTG CCA TGA AAA -3'**  
 31 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCT TAG AGC CTC CCA GTG AAA -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GGA CAA AGG CCC GGT CTC -3'**  
 32 **F 5'- CGT ATC GCC TCC CTC GCG CCA TCA GCT CGC ATC TCT CCT TCA CG -3'**  
**R 5'- CTA TGC GCC TTG CCA GCC CGC TCA GAG CCA ACG GAG TCG C -3'**

Supplementary table 2:

Physico-chemical and post translational modifications analysis of HTLV-1 Cosmopolitan

Transcontinental sequences with clinical forms data available on GenBank database.

Gene (protein)	Physico-chemical analysis	Post translational modifications	Clinical Profile			p-value
			ASY	TSP/HAM	ATL	
<b>Gag (p15)</b>			n=36	n=28	n=5	
P8S	-	Creation cAMP- and cGMP- dependent protein kinase phosphorylation site	-	-	-	-
P77R	-	Creation cAMP- and cGMP- dependent protein kinase phosphorylation site	-	-	-	-
<b>Gag (p19)</b>			n=36	n=28	n=5	
T106I	Reduced antigenicity	-	-	-	-	-
<b>Gag (p24)</b>			n=36	n=28	n=5	
I153V	Increased antigenicity	-	-	-	-	-
<b>Pro</b>			n=36	n=28	n=5	
G201D	-	Abrogation N-myristoylation site	4	3	1	0.795
<b>Pol</b>			n=36	n=28	n=5	
T40A	-	Abrogation protein kinase C phosphorylation site	-	-	-	-
Q446R	-	Creation protein kinase C phosphorylation site	13	12	2	0.859
<b>Env (gp46)</b>			n=38	n=34	n=6	
W88S	-	Creation protein kinase C phosphorylation site	-	-	-	-
P92S	-	Creation protein kinase C phosphorylation site	-	-	-	-
V247I	Reduced antigenicity	-	9	10	-	0.008
G307R	-	Creation protein kinase C phosphorylation site	2	-	-	0.339
<b>pX (Tax)</b>			n=37	n=30	n=6	
D264G	-	Creation N-myristoylation site	-	-	-	-
<b>pX (p13)</b>			n=36	n=30	n=5	
R47C	-	Abrogation protein kinase C phosphorylation site	-	-	-	-
<b>pX (p30)</b>			n=36	n=28	n=5	
R37C	-	Abrogation protein kinase C phosphorylation site	14	12	-	0.185

Position related to J02029. Asy, asymptomatic; TSP/HAM, Tropical Spastic Paraparesis /HTLV-1-Associated Myelopathy;

ATL, Adult T-cell leukemia/lymphoma. P- values were calculated by the Fisher's exact test

## 6 DISCUSSÃO

A partir do desenvolvimento da primeira etapa deste estudo foi possível constatar a escassez de dados sobre as sequências disponíveis publicamente. Informações tais como gênero, idade, perfil clínico, carga proviral, data de coleta e localização geográfica são de suma importância para os estudos moleculares e epidemiológicos do HTLV-1. Considerando estas dificuldades, ressalta-se a importância do fornecimento e disponibilidade pública dos dados gerados e coletados sobre as sequências virais, respeitando a propriedade intelectual, quando estes forem factíveis para o conhecimento científico sobre o HTLV-1.

Da construção de *softwares* relacionados ao HTLV-1 à utilização de sequências para as diversas análises de Bioinformática perpassa a necessidade de uma mineração de dados eficiente e robusta. No capítulo do livro “*Molecular epidemiology database for HTLV-1 sequence management, data mining and genotyping*” objetivou-se a construção de um protocolo didático, que auxilia na obtenção e gerenciamento de sequências e de informações relacionadas a elas, acrescido à possibilidade de uma rápida subtipagem viral a partir da região LTR do HTLV-1. O *HTLV-1 Molecular Epidemiology Database* (ARAÚJO et al., 2012) e *HTLV-1 Subtyping Tool* (ALCANTARA et al., 2009) são ferramentas que favorecem as pesquisas relacionadas ao HTLV-1, principalmente aquelas ligadas à infecção, patogênese, origem e evolução do vírus.

Na 15ª Conferência Internacional em Retrovirologia humana realizada na Bélgica em 2011, a maioria dos pesquisadores com expertise na infecção pelo HTLV-1 levantaram a necessidade de maiores estudos envolvendo a caracterização do genoma completo do vírus, com o intuito de favorecer as

pesquisas relacionados à patogenicidade e à compreensão da história natural em hospedeiros humanos e não humanos (MARTIN et al., 2011). Levando-se isto em consideração, iniciamos a terceira etapa do projeto que consiste na geração e caracterização de genomas completos provenientes de indivíduos com diferentes perfis clínicos.

Do primeiro genoma completo do HTLV-1 gerado em 1983 (SEIKI et al., 1983) até os dias atuais (2016), cerca de 91 genomas completos estão disponíveis, uma média de 0,12 genomas completos por 10.000 indivíduos infectados em todo o mundo, uma quantidade muito aquém em relação a um outro importante retrovírus humano, o HIV-1, que até as últimas atualizações do *GenBank*, possui quase dois mil genomas completos disponíveis (média de 0,55 genomas completos por 10.000 indivíduos infectados). Esta expressiva diferença pode ser reflexo do número reduzido de investimento em pesquisas relacionados ao HTLV-1, haja vista que, exceto o Japão, a grande parte dos países que são endêmicos para esta infecção, ou são subdesenvolvidos ou não desenvolvidos. Assim, estudos classificam a infecção pelo HTLV-1 e suas doenças associadas como uma infecção emergente, porém completamente negligenciada (CASSEB, 2009; GONÇALVES et al., 2010; COSTA et al., 2013).

Atualmente com o desenvolvimento deste trabalho, 31 genomas completos do HTLV-1, provenientes de indivíduos com TSP/HAM, DIH, ATL e assintomáticos foram gerados e analisados. Ressalta-se que os genomas oriundos de indivíduos com DIH serão os primeiros disponibilizados publicamente. Estas sequências foram geradas a partir de uma tecnologia de sequenciamento de alta performance, caracterizada por uma grande eficiência e baixo custo, tornando-se mais acessível para a realização de estudos



envolvendo a diversidade genômica viral (RADFORD et al., 2012; CAPOBIANCHI et al., 2013).

Em relação à análise filogenética utilizando a região LTR e o genoma completo, nossos resultados corroboraram com os estudos anteriores, confirmando o subtipo a (Cosmopolita), subgrupo A (Transcontinental) como o mais prevalente no país (DOURADO et al., 2003; ALCANTARA et al., 2006; CARNEIRO-PROIETTI et al., 2012; PESSÔA et al., 2014).

As distâncias genéticas intragrupo e intergrupo demonstraram baixa diversidade. Esta característica do genoma do HTLV-1 permite utilizá-lo em estudos específicos envolvendo eventos de migrações das populações humanas, na elucidação dos mecanismos da evolução viral e monitoramento dos meios de transmissão (GESSAIN et al., 1992).

No que concerne à relação entre variantes e perfis clínicos nas 31 sequências geradas, os dados deste estudo não tiveram suporte estatístico. O mesmo resultado foi encontrado quando se ampliou as análises para as sequências de HTLV-1aA com o perfil clínico definido, disponíveis no banco de dados *GenBank* e no *HTLV-1 Molecular Epidemiology Database*. A maioria das divergências nucleotídicas entre os isolados é caracterizada por mutações pontuais, que por sua vez, não estariam relacionadas aos diferentes perfis clínicos associados ao HTLV-1. Estas mutações seriam reflexo da origem geográfica e/ou da migração de populações antigas com indivíduos infectados pelo HTLV-1 (KOMURIAM et al., 1991; FERREIRA et al., 1997; SEGURADO et al., 2002).

Na análise da região promotora viral (*LTR*), mutações foram capazes de provocar a perda do sítio de ligação do fator de transcrição *Sp1* em *TxREs* específicos. O fator *Sp1* é crítico para a transcrição do *HTLV-1 bZIP factor (HBZ)*, que desempenha um papel importante na proliferação de células infectadas (YOSHIDA et al., 2008).

Além disso, foram identificadas duas mutações na região codificante que foram capazes de reduzir a antigenicidade e uma terceira mutação que aumentou a antigenicidade. Não podemos associar estes resultados encontrados com o perfil clínico, devido ao número reduzido de indivíduos recrutados (31 sequências) neste estudo, além da falta de associação estatística considerando outros genomas disponíveis publicamente. No entanto, estas mutações podem se tornar biomarcadores que estão localizados em domínios de proteínas importantes.

Os eventos de modificação pós-traducionais desempenham um papel importante na replicação viral e transformação celular. Nossas análises demonstraram uma alta identidade entre as sequências, contendo tais sítios, o que sugere a influência desses locais na resposta imune do hospedeiro e persistência do vírus. Identificamos uma baixa frequência de mutações associadas com a criação ou a revogação desses sítios em diferentes perfis clínicos. Algumas modificações como fosforilação, ubiquitinação e acetilação são críticos para transativação de *Tax* via CREB/ATF e NF- $\kappa$ B, que atua na inibição da reparação do DNA, no controle do ciclo celular e ativação de supressor de tumor p53 (JEANG et al., 2001; PELOPONESE et al., 2004; LAMSOUI et al., 2005; DURKIN et al., 2006; NASR et al., 2006; LODEWICK et al., 2009; LODEWICK et al., 2011; KFOURY et al., 2012). As modificações pós-tradução

de outras proteínas não estão bem documentados, por conseguinte, existe uma falta de evidência da importância destes locais para infecção viral (KANNAGI et al., 1992).

Além dos dados suscitados neste trabalho, os genomas completos do HTLV-1 e as mutações encontradas podem ser utilizadas para diversas avaliações, inclusive acerca da estabilidade do genoma e nas suas características estruturais, por exemplo, predizendo as suas estruturas secundárias. Já iniciamos esta avaliação, e dados preliminares apontam que a baixa variabilidade genômica do HTLV-1 repercute na manutenção das estruturas secundárias do genoma.

## 7 CONSIDERAÇÕES FINAIS

- A partir da avaliação das variáveis moleculares, clínicas e epidemiológicas dos dados depositados no *HTLV-1 Molecular Epidemiology Database*, constatamos uma expressiva escassez de dados sobre as sequências disponíveis, e conseqüentemente, um limitador para os estudos envolvendo a diversidade viral, estudos clínicos e epidemiológicos.
- A publicação do capítulo que sistematiza a mineração de dados e a subtipagem do HTLV-1 torna-se um instrumento importante de ensino/aprendizagem relacionadas à utilização das ferramentas: *HTLV-1 Subtype tool* e *HTLV-1 Molecular Epidemiology Database*. Estas ferramentas são importantes nos estudos relacionados à epidemiologia molecular e genotipagem do HTLV-1.
- Este estudo contribui para o aumento de genomas completos provenientes de pacientes infectados pelo HTLV-1 com diferentes perfis clínicos: indivíduos com TSP/HAM, ATL, DIH e assintomáticos.
- Os genomas completos apresentaram uma baixa diversidade intra e intergrupo, além disso, não tiveram associação estatística entre as mutações e os perfis clínicos analisados.
- As alterações do perfil físico-químico, o impacto das mutações em relação à afinidade de ligação dos fatores de transcrição na região *LTR* e as análises dos sítios de modificações pós-traducionais não tiveram associação estatística em relação aos diferentes perfis clínicos, considerando as sequências geradas neste trabalho e nas previamente disponíveis no *GenBank*.
- Sugerimos que para uma melhor investigação da contribuição das mutações do HTLV-1 com os possíveis desfechos clínicos, faz-se necessário avaliar a interação do genoma viral e as características do hospedeiro humano.

## REFERÊNCIAS

ALCANTARA L. C. J. et al. Tracing the origin of brazilian HTLV-1 as determined by analysis of host and viral genes. **AIDS**. v. 20, p. 780-782, 2006.

ALCANTARA L. C. J et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. **Nucleic Acids Research**. v. 37, n. 9, p. 634-642, 2009.

ARAÚJO, A. Q. C; SILVA, M. T. T. The HTLV-1 neurological complex. **Lancet Neurological**. v. 5, n. 12, p. 1068–1076, 2005.

ARAÚJO T. H. A. et al. A public HTLV-1 molecular epidemiology database for sequence management and data mining. **Plos One**. v. 7, n. 9, p. 1-10, 2012.

ARAÚJO, T. H. A. et al. M. Inferences about the global scenario of human T-cell lymphotropic virus type 1 infection using data mining of viral sequences. **Memórias do Instituto Oswaldo Cruz**. v. 109, n. 4, p. 448-451, 2014.

BAGOSSI P. et al. Discovery and significance of new human T-lymphotropic viruses: HTLV-3 and HTLV-4. **Expert Review of Anti-infective Therapy**. v. 7, p. 1235-1249, 2009.

BANDEIRA, L. M. et al. High prevalence of HTLV-1 Infection among japanese immigrants in non-endemic area of Brazil. **PLoS Neglected Tropical Diseases**. v. 9, n. 4, p. 1-11, 2015.

BANGHAM, C.R. The immune response to HTLV-1. **Current Opinion in Immunology**. v.12, p. 397-402, 2000.

BANGHAM, C. R. M. The immune control and cell-to-cell spread of human T-lymphotropic virus type 1. **Journal of General Virology**. v. 84, n. 12, p. 3177-3189, 2003.

BANGHAM, C. R.; OSAME, M. Cellular immune response to HTLV-1. **Oncogene**. v. 24, p. 6035-6046, 2005.

BASTIAN I., J. et al. Isolation of a human T-lymphotropic virus type I strain from Australian aboriginals. **Journal of Virology**. v. 67, p. 843-851, 1993.

BATTINI, J. L. et al. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. **Cell**. v. 115, p. 449-459, 2003.

BISWAS, H. H. et al. Neurologic abnormalities in HTLV-I- and HTLV-II infected individuals without overt myelopathy. **Neurology**. v. 73, p. 781-789, 2009.

BITTENCOURT, A. L. Vertical Transmission of HTLV-I/II: a review. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 40, n. 4, p. 245-251, 1998.

BITTENCOURT, A. L.; OLIVEIRA, M. F. Infective dermatitis associated with the HTLV-I (IDH) in children and adults. **Anais Brasileiros de Dermatologia**. v. 80, p. 364-369, 2005.

BITTENCOURT, A. L et al. Adult T-cell leukemia/lymphoma in Bahia, Brazil. Analysis of prognostic factors in a group of 70 patients. **American Journal of Clinical Pathology**. v. 128, p. 875-882, 2007.

BITTENCOURT, A. L.; FARRÉ, L. Leucemia/linfoma de células T do adulto/ Adult T-cell leukemia/lymphoma. **Anais Brasileiros de Dermatologia**. v. 83, n. 4, p. 351-9, 2008.

BURKE, D. S. Recombination in HIV: An important evolutionary strategy. **Emergent Infectious Diseases**, v. 3, p. 253-258, 1997.

CALATTINI, S. et al. A New strain of human T lymphotropic virus (HTLV) type 3 in a Pygmy from Cameroon with peculiar HTLV serologic results. **Journal of Infectious Diseases**. v. 199, p. 561-564, 2009. CALATTINI, S. et al. Discovery of a new human T-cell lymphotropic virus (HTLV-3) in Central Africa. **Retrovirology**. v. 2, p. 1-4, 2005.

CANN A. J; CHEN S. Y. Human T-cell leukemia virus types I and II. **Fields Virology**. 3 ed. Philadelphia: Raven Publishers. v. 2, p. 1849-1879, 1996.

CARNEIRO-PROIETTI A. B. F. et al. Infecção e doença pelos vírus linfotrópicos humanos de células T (HTLV-I/II) no Brasil. **Revista Brasileira da Sociedade de Medicina Tropical**. v. 35, p. 499-508, 2002.

CARRAZZONE, C. F. V.; BRITO, A. M; GOMES, Y. M. The importance of pre-transfusional serological screening in blood transfusion recipients. **Revista Brasileira de Hematologia e Hemoterapia**. v. 26, n. 2, p. 93-98, 2004.

CASSAR, O. et al. Human T-cell lymphotropic virus type 1 subtype c molecular variants among indigenous australians: new insights into the molecular epidemiology of HTLV-1 in Australo-Melanesia. **PLoS Neglected Tropical Diseases**. v. 7, n. 9, p. 1-11, 2013.

CASSEB, J. Is human T-cell lymphotropic type 1 (HTLV-1) associated myelopathy/tropical spastic paraparesis (HAM/TSP): syndrome a neglected disease? **PLoS Neglected Tropical Diseases**. v. 3, p. 1-3, 2009.

CATALAN-SOARES, B. C. et al. HLA class I alleles in HTLV-1- associated myelopathy and asymptomatic carriers from the brazilian cohort GIPH. **Medical Microbiology and Immunology**. v. 198, p. 1-3, 2009.

CATALAN-SOARES, B.; CARNEIRO-PROIETTI, A. B.; PROIETTI F. A. Vírus-T linfotrópico humano em familiares de candidatos a doação de sangue soropositivos: disseminação silenciosa. **Revista Panamericana Salud Pública/Pan American Journal Public Health**. v. 16, n. 6, p. 387-394, 2004.

CATALAN-SOARES, B.; CARNEIRO-PROIETTI, A. B.; PROIETTI, F. A. Heterogeneous geographic distribution of human T-cell lymphotropic viruses I and II (HTLV-I/II): serological screening prevalence rates in blood donors from large urban areas in Brazil. **Caderno de Saúde Pública**. v. 21, p. 926-931, 2005.

CHEN, J. L. et al. HTLV isolated from a Pygmy in Cameroon is related but distinct from the known Central African type. **AIDS Research and Human Retroviruses**. v. 11, p. 1529-1531, 1995.

CIMARELLI, A. et al. Clonal expansion of human T-cell leukemia virus type II in patients with high proviral load. **Virology**. v. 223, p. 362-364, 1996.

CONSTATINE, N. T.; CALLAHAN, J. D.; WATTS, D. M. Retroviral testing. essential for quality control and laboratory diagnosis. **CRC Press**. p. 105-107, 1992.

COSKUN, A. K.; SUTTON, R. E. Expression of glucose transporter 1 confers susceptibility to human T-cell leukemia virus envelope-mediated fusion. **Journal Virology**. v. 79, n. 7, p. 4150-4158, 2005.

COSTA, C.A. et al. Familial transmission of human T-cell lymphotropic virus: silent dissemination of an emerging but neglected infection. **PLOS Neglected Tropical Diseases**. v. 7, n. 6, p. 1-7, 2013.

DAENKE, S. et al. Sequence variants of human T-cell lymphotropic virus type I from patients with tropical spastic paraparesis and adult T-cell leukemia do not distinguish neurological from leukemic isolates. **Journal Virology**. v. 64, p. 1278-1282, 1990.

DERSE, D. et al. Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. **Proceedings of the National Academy of Sciences**. v. 104, n. 8, p. 2915–2920, 2007.

DESRAMES, A. et al. Northern African strains of human T-lymphotropic virus type 1 arose from a recombination event. **Journal Virology**. v. 88, p. 9782–9788, 2014.

DE THÉ, G.; KAZANJI, M. An HTLV-I/II vaccine: from animal models to clinical trials? **Journal Acquired Immune Deficiency Syndromes and Human Retrovirology**. v. 13, p. 191-198, 1996.

DOONEIEF, G. et al. Neurologic consequences of HTLV-II infection in injectiondrug users. **Neurology**. v. 46, p. 1556-1560, 1996.

DOURADO, I. et al. HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. **Journal Acquired Immune Deficiency Syndromes and Human Retrovirology**. v. 34, n. 5, p. 527-531, 2003.

DRUMMOND, A. J.; RAMBAUT, A. **BEAST v. 1.0**. Disponível em <http://evolve.zoo.ox.ac.uk/beast/>. Acesso em 10 de setembro de 2016.

EDWARDS, D. et al. HTLV-1 orf-I and orf-II-encoded proteins in viral infection and persistence. **Viruses**. v. 3, n. 6, p. 861- 885, 2011.

ENOSE-AKAHATA, Y. et al. Development of neurologic diseases in a patient with primate T-lymphotropic virus type 1 (PTLV-1). **Retrovirology**. v. 13, n. 56, p. 1-13, 2016.

EVANGELISTA, A. et al. Nucleotide sequence analysis of a provirus derived from an individual with tropical spastic paraparesis. **Microbial Pathogenesis**. v. 8, n. 4, p. 259-78, 1990.

FERREIRA, O. C.; PLANELLES, V.; ROSENBLATT, J. D. Human T-cell leukemia viruses: epidemiology, biology, and pathogenesis. **Blood Reviews**. v.11, p. 91-104, 1997.

FEUER, G.; GREEN, P. L. Comparative biology of human T-cell lymphotropic virus type 1 (HTLV-1) and HTLV-2. **Oncogene**. v. 24, p. 5996-6004, 2005.

FURUKAWA, Y. et al. Different cytokine production in tax-expressing cells between patients with human T-cell lymphotropic virus type I (HTLV-I)-associated myelopathy/tropical spastic paraparesis and asymptomatic HTLV-I carriers. **Journal of Infectious Diseases**. v. 187, p. 1116-1125, 2003.

FURUYA, T. et al. HTLV-I associated myelopathy associated with multi-organ inflammatory disease: a case report. **Journal of the Neurological Sciences**. v. 157, p. 109-112, 1998.

GALLO, R. C. Kyoto Workshop on some specific recent advances in human tumor virology. **Cancer Research**. v. 41, p. 4738-4739, 1981.

GASMI, M.; FAROUQI, M.; D'INCAN, C. Desgranges long terminal repeat sequence analysis of HTLV type I molecular variants identified in four north African patients. **AIDS Research and Human Retroviruses**. v. 10, p. 1313-1315, 1994.

GESSAIN, A. et al. Antibodies to human T- lymphotropic virus type-I in patients with tropical spastic paraparesis. **Lancet**. v. 2, p. 407-410, 1985.

GESSAIN, A. et al. Highly divergent molecular variants of human T-lymphotropic virus type from isolated populations in Papua New Guinea and the Solomon Islands. **Proceedings of the National Academy of Sciences**. v. 88, p. 7694-7698, 1991.

GESSAIN, A; GALLO, RC; DE THÉ, G. Genetic variability and molecular epidemiology of Human and Simian T-cell leukemia/ lymphoma virus type 1. **Journal Acquired Immune Deficiency Syndromes and Human Retrovirology**. v. 13, p. 132-145, 1992.



GESSAIN, A, Gout O. Chronic myelopathy associated with human T-lymphotropic virus type I (HTLV-I). **Annals of Internal Medicine.** v. 117, p. 933-946, 1992.

GESSAIN, A.; DE THE, G. Geographic and molecular epidemiology of primate T lymphotropic retroviruses: HTLV-I, HTLV- II, STLV-I, STLVPP and PTLV-L. **Journal Acquired Immune Deficiency Syndromes and Human Retrovirology.** v. 13, p. 228-235, 1996.

GESSAIN, A.; CASSAR, O. Epidemiological Aspects and World Distribution of HTLV-1. **Frontiers in Microbiology.** v. 3, p. 1-23, 2012.

GHEZ, D. et al. Neuropilin-1 is involved in human T-cell lymphotropic virus type 1. **Journal of Virology.** v. 80, n. 14, p. 6844-6854, 2006.

GONÇALVES, D. U. et al. Dermatologic lesions in asymptomatic blood donors seropositive for Human T-cell lymphotropic virus type-1. **American Journal of Tropical Medicine and Hygiene.** v. 68, p. 562-565, 2003.

GONÇALVES, D. U. et al. Epidemiology, treatment and prevention of human T-cell leukemia virus type 1 - associated diseases. **Clinical Microbiology Reviews.** v. 23, n. 3, p. 577-589, 2010.

GRASSI, M. F. et al. Utility of HTLV proviral load quantification in diagnosis of HTLV-1-associated myelopathy requires international standardization. **Journal of Clinical Virology.** v. 58, n. 3, p. 1-5, 2013.

GREEN, P. L; CHEN, S. Y. Human T-cell leukemia virus types 1 and 2. **Fields Virology.** 4 ed., Lippincott Williams & Wilkins, Philadelphia, 2001.

GUANGYONG, M. et al. TCF1 and LEF1 act as T-cell intrinsic HTLV-1 antagonists by targeting Tax. **PNAS.** v. 112, n. 7, p. 2216–2221, 2015.

HAHN, B. H, et al. Molecular cloning and analysis of a new variant of human T-cell leukemia virus (HTLV-Ib) from an African patient with adult T-cell leukemia-lymphoma. **International Journal of Cancer.** v. 34, n. 5, p. 613-618, 1984.

HADDAD, R. et al. HLA-G 14-bp insertion/deletion polymorphism is a risk factor for HTLV-1 infection. **AIDS Res Hum Retroviruses.** v. 27, p. 283-288, 2011.

HANCHARD, B. et al. **Humman Retrovirology: HTLV.** Blattner WA ed). Raven Press: New York, p. 173-183,1990.

HANON, E., et al. Fratricide among CD8(+) T lymphocytes naturally infected with human T cell lymphotropic virus type I. **Immunity.** v. 13, n. 5, p. 657-664, 2000.

HÖLLSBERG, P.; HAFLER, D. A. What is the pathogenesis of human T-cell lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis? **Annals Neurology.** v. 37, p. 143-145, 1995.

HJELLE, B. et al. Chronic neurodegenerative disease associated with HTLV-II infection. **Lancet**. v. 339, p. 645-646, 1992.

JEANMOUGIN, F. et al. Multiple sequence alignment with clustal X. **Trends in Biochemical Sciences**. v. 23, p. 403-405, 1998.

JONES, K. S., et al. Molecular aspects of HTLV-1 entry: functional domains of the HTLV-1 surface subunit (SU) and their relationships to the entry receptors. **Viruses**. v. 3, n. 12, p. 794-810, 2011.

KALYANARAMAN, V. S. et al. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. **Science**. v. 218, n. 4572, p. 571-573, 1982.

KASHIMA, S. et al. Distribution of human T cell lymphotropic virus type 1 (HTLV-1) subtypes in Brazil: genetic characterization of LTR and tax region. **AIDS Research and Human Retroviruses**. v. 22, n. 10, p. 953-959, 2006.

KNIGHT, S. C. et al. Dendritic cells in HIV-1 and HTLV-1 infection. **Advances in Experimental Medicine and Biology**. v. 329, p. 545-549, 1993.

KOYANAGI, Y. et al. In vivo infection of human T-cell leukemia virus type I in non-T cells. **Virology**. v. 196, p. 25-33, 1993.

KOMURIAN, F.; PELLOQUIN, F.; DE THE, G. In vivo genomic variability of human T-cell leukemia virus type I depends more upon geography than upon pathologies. **Journal Virology**. v. 65, p. 3770–3778, 1991.

LA GRENADE, L. et al. Infective dermatitis of Jamaican children: a marker for HTLV-I infection. **Lancet**. v. 336, p. 1345-1347, 1990

LA GRENADE, L. et al. Clinical, pathologic and immunologic features of human T-lymphotrophic virus type I-associated infective dermatitis in children. **Archives of Dermatology**. v. 134, p. 439-444, 1998.

LA GRENADE, L. HTLV-I-associated infective dermatitis: past, present, and future. **Journal of Acquired Immune Deficiency Syndromes**. v. 13, p. 42-49, 1996.

LAL, R. B. et al. Sensitivity and specificity of a recombinant transmembrane glycoprotein (rgp21)-spiked western immunoblot for serological confirmation of human T-cell lymphotropic virus type I and type II infections. **Journal of Clinical Microbiology**. v. 30, n. 2, p. 296-299, 1992.

LAMB, D. et al. Highly specific inhibition of leukaemia virus membrane fusion by interaction of peptide antagonists with a conserved region of the coiled coil of envelope. **Retrovirology**. v. 5, p. 1-14, 2008.

LEMEY, P. et al. A Bayesian statistical analysis of human T-cell lymphotropic virus evolutionary rates. **Infection, Genetics and Evolution**. v. 5, p. 291-298, 2005.

LEE, S. M. et al. HTLV-1 induced molecular mimicry in neurological disease. **Current Topics in Microbiology and Immunology**. v. 296, p. 125-36, 2005.

LEPOUTRE, V. et al. Role of resident CNS cell populations in HTLV-1-associated neuroinflammatory disease. **Frontiers in Bioscience**. v. 14, p. 1152-1168, 2009.

LEVIN, M. C; JACOBSON, S. Cellular and humoral immune responses associated with HTLV-I-associated myelopathy/tropical spastic paraparesis. **Annals of the New York Academy of Sciences**. v. 835, p. 142-52, 1997.

LIU, H.F. et al. The three human T-lymphotropic virus type I subtypes arose from three geographically distinct simian reservoirs. **Journal of General Virology**. v. 77, p. 359–368, 1996.

KOMURIAN, F.; PELLOQUIN, F.; DE THÉ, G. In Vivo genomic variability of human T-cell leukemia virus type I depends more upon geography than upon pathologies. **Journal of Virology**. v. 65, n. 7, p. 3770-3778, 1991.

KUBOTA, R. et al. Degenerate specificity of HTLV-I specific CD8+ T cells during viral replication in patients with HTLV-1-associated myelopathy (HAM/TSP). **Blood**. v. 101, p. 3074-81, 2003.

KURAMITSU, M. et al. Identification of TL-Om1, an adult T-cell leukemia (ATL) cell line, as reference material for quantitative PCR for human T lymphotropic virus 1. **Journal of Clinical Microbiology**. v. 53, n. 2, p. 587-596, 2015.

IWANAGA, M. et al. Human T-cell leukemia virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan. **Blood**. v. 116, n. 8, p. 1211-9, 2010.

IJICHI, S. et al. HAM/TSP: recent perspectives in Japan. **Journal of Acquired Immune Deficiency Syndromes**. v. 13, p. 26-32, 1996.

IJICHI, S.; OSAME, M. Human T lymphotropic virus type I (HTLV-I)- associated myelopathy/tropical spastic paraparesis (HAM/TSP): recent perspectives. **Journal of Internal Medicine**. v. 34, p. 713-721, 1995.

MA, Y. et al. Epidemiological Analysis of HTLV-1 and HTLV-2 Infection among Different Population in Central China. **PLoS ONE**. v. 8, n. 6, p. 1-8, 2013.

MAHIEUX, R. et al. Molecular epidemiology of 58 new African human T-cell leukemia virus type 1 (HTLV-1) strains: identification of a new and distinct HTLV-1 molecular subtype in Central Africa and in Pygmies. **Journal of Virology**. v. 71, n. 2, p. 1317-33, 1997.

MALIK, K.T.; EVEN, J.; KARPAS, A. Molecular cloning and complete nucleotide sequence of an adult T cell leukaemia virus/human T cell leukaemia virus type I (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/HTLV-I subgroup. **Journal of General Virology**. v. 69, p. 1695-1710, 1988.

MARTINS, M. L.; STANCIOLI, E. F. B. Vírus linfotrópico de células T humanas Tipos 1 e 2 (HTLV – 1/2) – patogênese da infecção pelo HTLV. **Cadernos Hemominas. HTLV**. Volume XIII, 4<sup>a</sup> ed., Belo Horizonte, p. 21-45, 2006.

MARTIN, F. et al. Conference highlights of the 15th international conference on human retrovirology: HTLV and related retroviruses, Leuven, Gembloux, Belgium. **Retrovirology**. v. 8, n. 86, p. 1-15, 2011.

MAJOROVITS, E. et al. Human T-lymphotropic virus-1 visualized at the virological synapse by electron tomography. **PLoS ONE**. v. 3, p. 1-10, 2008.

MALONEY, E. M. et al. Human T cell lymphotropic virus type I-associated infective dermatitis in Jamaica: A case report of clinical and biologic correlates. **Pediatric Infectious Disease Journal**. v. 19, p. 456-461, 2000.

MANEL, N. et al. HTLV envelopes and their receptor GLUT1, the ubiquitous glucose transporter: a new vision on HTLV infection? **Frontiers Bioscience**. v. 9, p. 3218-3241, 2004.

MANNS, A.; HISADA, M.; LA GRENADE, L. Human T-lymphotropic virus type I infection. **Lancet**. v. 353, p. 1951-1958, 1999.

MARDIS, E.R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**. v. 9, p. 387–402, 2008.

MATSUOKA, M. Human T-cell leukemia virus type I (HTLV-I) infection and the onset of adult T-cell leukemia (ATL). **Retrovirology**. v. 2, n. 27, p. 1-13, 2005.

MATSUOKA, M.; JEANG, K. T. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. **Nature**. v. 7, p. 270-280, 2007.

MCCALLUM, R. M. et al. Arthritis syndromes associated with human T cell lymphotropic virus type I infection. **Medical Clinics of North America**. v. 81, p. 261-276, 1997.

MIURA, T. et al. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. **Proceedings of the National Academy of Sciences**. v. 91, p. 1124-1127, 1994.

MIURA, T. et al. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. **Journal of Molecular Evolution**. v. 44, p. 76-82, 1997.

MOCHIZUKI, M. et al. Uveitis associated with human T lymphotropic virus type I: seroepidemiologic, clinical, and virologic studies. **Journal Infectious Diseases**. v. 166, p. 943–944, 1992.

MOCHIZUKI, M. et al. HTLV-I uveitis. **Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology**. v.13, p. 50-56, 1996.

MONTANHEIRO, P. A. et al. The elevated interferon gamma production is an important immunological marker in HAM/TSP pathogenesis. **Scandinavian Journal of Immunology**. v. 70, n. 4, p. 403-497, 2009.

MORGAN, O. S. et al. HTLV-I and polymyositis in Jamaica. **Lancet**. v. 2, p. 1184-1187, 1989.

MOSLEY, A. J.; ASQUITH, B.; BANGHAM, C. R. Cell-mediated immune response to human T lymphotropic virus type I. **Viral Immunology**. v. 18, p. 293-305, 2005.

MUELLER, N. The epidemiology of HTLV-I infection. **Cancer Causes Control**. v. 2, n.1, p. 37-52, 1991.

MUELLER, N. et al. Findings from the Miyazaki cohort study. **Journal of Acquired Immune Deficiency Syndromes**. v. 13, p. 1-7, 1996.

MURPHY, E. L. The Clinical Epidemiology of Human T-Lymphotropic Virus Type II (HTLV-II). **Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology**. v. 13, n. 1, p. 215-219, 1996.

NAKADA, K. et al. High incidence of HTLV antibody in carriers of *Strongyloides stercoralis*. **Lancet**. v. 1, p. 633, 1984.

NAKAMURA, H. et al. Relationship between Sjögren's syndrome and human T-lymphotropic virus type I infection: follow-up study of 83 patients. **Journal of Laboratory and Clinical Medicine**. v. 135, p. 139-44, 2000.

NATH, M. D. Regulation of the cell-surface expression of an HTLV-I binding protein in human T cells during immune activation. **Blood**. v. 101, n. 8, p. 3085-3092, 2002.

NEJMEDDINE, M.; BANGHAM, C. R. M. The HTLV-1 virological synapse. **Viruses**. v. 2, p. 1427-1447, 2010.

NISHIOKA, K. HTLV-1 arthropathy and Sjögren syndrome. **Retrovirology**. v. 13, p. 57-62, 1996.

OSAME, M. et al. HTLV-I associated myelopathy, a new clinical entity. **Lancet** v. 1, p. 1031-1032, 1986.

OSAME, M. Pathological mechanisms of human T-cell lymphotropic virus type I-associated myelopathy (HAM/TSP). **J Neurovirol**. v. 8, p. 359-64, 2002.

PAIS-CORREIA, A. M. et al. Biofilm-like extracellular viral assemblies mediate HTLV-1 cell-to-cell transmission at virological synapses. **Nat Med**. v. 16, n. 1, p. 83-90, 2009.

PESSÔA, R. et al. Molecular Characterization of Human T-Cell Lymphotropic Virus Type 1 Full and Partial Genomes by Illumina Massively Parallel Sequencing Technology. **Plos One**. v.9, n.3, 2014.

PINON, J. D., et al. Human T-Cell Leukemia Virus Type 1 Envelope Glycoprotein gp46 Interacts with Cell Surface Heparan Sulfate Proteoglycans. **Journal of Virology**. v. 77. n. 18. p. 9922-9930. 2003.

POIESZ, B. J. et al. 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**. v. 77, p. 7415-7419, 1980.

POND, S. L.; FROST, S. D.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. **Bioinformatics**. v. 21, n. 5, p. 676–679, 2005.

PORTO, A. F. et al. Influence of human T-cell lymphocytotropic virus type 1 infection on serologic and skin tests for strongyloidiasis. **American Journal of Tropical Medicine and Hygiene**. v. 65, n. 5, p. 610-613, 2001.

PROIETTI, F. A. et al. Global epidemiology of HTLV-1 infection and associated disease. **Oncogene**. v. 24, p. 6058-6068, 2005.

PROIETT, A. B. F. C. **Cadernos Hemominas - HTLV** (2015). 6<sup>a</sup> ed., v. XVI. Belo Horizonte: FUNDAÇÃO HEMOMINAS.

RAMMENSEE, H. G. et al. SYFPEITHI: database for MHC ligands and peptide motifs. **Immunogenetics**. v. 50, p. 213–219, 1999.

RICHARDSON, J. H. et al. In vivo cellular tropism of human T-cell leukemia virus type 1. **Journal of Virology**. v. 64, p. 682-687, 1990.

ROUCOUX, D. F. et al. A prospective study of sexual transmission of human T lymphotropic virus (HTLV)-I and HTLV-II. **The Journal of infectious diseases**. v. 191, n. 9, p. 1490-1497, 2005.

RODGERS-JOHNSON, P. E. et al. Tropical spastic paraparesis and HTLV-I myelopathy: clinical features and pathogenesis. **Research Publications Association for Research in Nervous and Mental Disease**. v. 68, p. 117-130, 1990.

ROMÁN, G. C.; OSAME, M. Identity of HTLV-I-associated tropical spastic paraparesis and HTLV-I-associated myelopathy. **Lancet**. v.1, p. 651, 1988.

SABOURI, A. H. et al. Differences in viral and host genetic risk factors for development of human T-cell lymphotropic virus type 1 (HTLV-1)-associated

myelopathy/tropical spastic paraparesis between Iranian and Japanese HTLV-1-infected individuals. **Journal of General Virology**. v. 86, p. 773-781, 2005.

SALEMI, M. et al. Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have closest relatives among African STLV-I strains. **Virology**. v. 246, p. 277-287, 1998.

SAKAI, J. A. et al. In vitro spontaneous lymphoproliferation in patients with human T-cell lymphotropic virus type I-associated neurologic disease: predominant expansion of CD8+ T cells. **Blood**. v. 98, p. 1506-1511, 2001.

SANTOS, F. L. N.; LIMA, F. W. M. Epidemiology, physiopathogenesis and laboratorial diagnosis of the HTLV-I infection. **Jornal Brasileiro de Patologia e Medicina Laboratorial**. v. 41, p. 105-16, 2005.

SATO, T. et al. CSF CXCL10, CXCL9, and neopterin as candidate prognostic biomarkers for HTLV-1-associated myelopathy/tropical spastic paraparesis. **PLoS Neglected Tropical Diseases**. v. 7, n. 10, p. 1-11, 2013.

SEIKI, M.; HATTORI, S.; YOSHIDA, M. Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. **Proceedings of the National Academy of Sciences**. v. 79, p. 6899-6902, 1982.

SEIKI, M. et al. 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**. v. 80, p. 3618–3622, 1983.

SEIKI, M. et al. Nonspecific integration of the HTLV provirus genome into Adult t-cell leukemia cells. **Nature**. v. 309, p. 640-642, 1984.

SEGURADO, A. A. et al. Identification of human T-lymphotropic virus type I (HTLV-I) subtypes using restricted fragment length polymorphism in a cohort of asymptomatic carriers and patients with HTLV-I-associated myelopathy/tropical spastic paraparesis from Sao Paulo, Brazil. **Memórias do Instituto Oswaldo Cruz**. v. 97, p. 329–333, 2002.

SETOYAMA, M. et al. Detection of HTLV-1 by polymerase chain reaction in situ hybridization in adult T-cell leukemia/lymphoma. **American Journal of Pathology**. v. 152, n. 3, p. 683-689, 1998.

SHERMAN, M. P. et al. Evolutionary insights on the origin of human T-cell lymphoma/leukemia virus type I (HTLV-I) derived from sequence analysis of a new HTLV-I variant from Papua New Guinea. **Journal of Virology**. v. 66, p. 2556-2563, 1992.

SHIMOTOHNO, K. et al. 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 Science**. v. 82, p. 3101–3105, 1985.

SHIMOYAMA, M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the lymphoma study group (1984-87). **British Journal of Haematology**. v. 79, p. 428-437, 1991.

SODROSKI, J. The human T-cell leukemia virus (HTLV) transactivator (Tax) protein. **Biochimica et Biophysica Acta**. v. 1114, p. 19-29, 1992.

SZE, A. et al. Host restriction factor SAMHD1 limits human T cell leukemia virus type 1 infection of monocytes via STING-mediated apoptosis. **Cell Host Microbe**. v. 14, n. 4, p. 422–434, 2013.

SWEET, R. D. A pattern of eczema in Jamaica. **British Journal of Dermatology**. v. 78, p. 93-100, 1966.

SWITZER, W. M. et al. Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. **Retrovirology**. v. 6, n. 9, p. 1-20, 2009.

SWOFFORD, D. **PAUP: phylogenetic analysis using parsimony**. Version 4.0b10. Smithsonian Institution, Washington, D.C, 1997.

THORSTENSSON, R.; ALBERT, J.; ANDERSON, R. Strategies for diagnosis of HTLV-I and -II. **Transfusion**. v. 42, p. 780-791, 2002.

TREVIÑO, A. et al. Molecular epidemiology and clinical features of human T cell lymphotropic virus type 1 infection in Spain. **AIDS Research and Human Retroviruses**. v. 30, p. 856-862, 2014.

TSCHACHLER, E.; FRANCHINI, G. Infective dermatitis a pabulum for human T-lymphotropic virus type I leukemogenesis? **Archives of Dermatological Research**. v. 134, n. 4, p. 487-488, 1998.

TSUJIMOTO, A. et al. Nucleotide sequence analysis of a provirus derived from HTLV-1 associated myelopathy (HAM). **Molecular Biology & Medicine**. v. 5, n. 1, p. 29-42, 1988.

URETA-VIDAL, A. et al. Human T-cell leukemia virus Type I (HTLV-I) infection induces greater expansions of CD8 T lymphocytes in persons with HTLV-I-associated myelopathy/tropical spastic paraparesis than in asymptomatic carriers. **Journal of Infectious Diseases**. v. 183, p. 857-864, 2001.

UCHIYAMA, T. et al. Adult T-cell leukemia: clinical and hematological features of 16 cases. **Blood**. v. 50, p. 481-491, 1977.

VANDAMME, A. M. et al. Primate T lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-I. **Virology**. v. 202, p. 212-223, 1994.



VAN DOOREN, S. et al. Evidence for a post-colombian introduction of human T-cell lymphotropic virus in Latin America. **Journal of General Virology**. v. 79, p. 2695-2708, 1998.

VAN PROOYEN, N., et al. Human T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission. **Proceedings of the National Academy of Sciences**. v. 107, n. 48, p. 20738-20743, 2010.

YAMAGUCHI, K. Human T-lymphotropic virus type I in Japan. **Lancet**. v. 343, p. 213-216, 1994.

YOSHIDA, M.; MIYOSHI, I.; HINUMA, Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. **Proceedings of the National Academy of Sciences**. v. 79, n. 6, p. 2031-2035, 1982.

WALSHE, M. M. Infective dermatitis in Jamaican children. **British Journal of Dermatology**. v. 79, p. 229-36, 1967.

WATTEL, E.; VARTANAIN, J. P.; PANNETRER, C. Clonal expansion of HTLV-I infected cells in asymptomatic and symptomatic carries without malignancy. **Journal of Virology**. v. 69, p. 2863–2868, 1995.

WOLFE, N. D. et al. Emergence of unique primate T lymphotropic viruses among central African bushmeat hunters. **Proceedings of the National Academy of Sciences**. v. 22, p. 7994-7999, 2005.

ZANELLA, L. et al. Genome-wide analyses of HTLV-1aD strains from Cape Verde, Africa. **Memórias do Instituto Oswaldo Cruz**. v. 111, n. 9, p. 594-596, 2016.

ZEHENDER, G. et al. Description of a trans-Saharan strain of human T-lymphotropic virus type 1 in West Africa. **Journal of Acquired Immune Deficiency Syndromes**. v.47, p. 269–273, 2008.

ZHI, H. et al. NF-kappaB hyperactivation by HTLV-1 Tax induces cellular senescence, but can be alleviated by the viral anti-sense protein HBZ. **PLoS Pathogens**. v. 7, p. 1-12, 2011.