

**FIOCRUZ**

**FUNDAÇÃO OSWALDO CRUZ  
CENTRO DE PESQUISAS GONÇALO MONIZ**

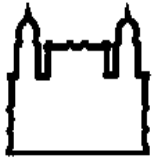
**CURSO DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA EM  
SAÚDE E MEDICINA INVESTIGATIVA**

**DISSERTAÇÃO DE MESTRADO**

**ESTUDO GENOTÍPICO DE NOVOS ISOLADOS DO  
HTLV-1 E CARACTERIZAÇÃO DO ENVELOPE  
VIRAL, CORRELACIONANDO COM O  
DESENVOLVIMENTO DE PROTÓTIPOS VACINAIS**

**ALINE CRISTINA ANDRADE MOTA MIRANDA**

**Salvador - Bahia – Brasil  
2008**



FIOCRUZ

**FUNDAÇÃO OSWALDO CRUZ  
CENTRO DE PESQUISAS GONÇALO MONIZ**

**CURSO DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA EM  
SAÚDE E MEDICINA INVESTIGATIVA**

**ALINE CRISTINA ANDRADE MOTA MIRANDA**

**ESTUDO GENOTÍPICO DE NOVOS ISOLADOS DO  
HTLV-1 E CARACTERIZAÇÃO DO ENVELOPE  
VIRAL, CORRELACIONANDO COM O  
DESENVOLVIMENTO DE PROTÓTIPOS VACINAIS**

Dissertação apresentada ao  
Curso de Pós-graduação em Biotecnologia  
em Saúde e Medicina Investigativa  
como requisito parcial para obtenção  
do título de Mestre.

**Orientador:** Prof. Dr. Luiz Carlos Júnior Alcântara

**Salvador - Bahia – Brasil  
2008**

"Ninguém acende uma lâmpada  
e a coloca num lugar onde ficará escondida,  
ou sob uma tigela..  
Ao invés disso, a coloca de pé,  
assim, aquele que entrar  
pode enxergar a luz."  
**Jesus Cristo**

Dedico este trabalho,  
e todo ensinamento e crescimento,  
conquistados pelo caminho,  
aos meus pais Vilson e Tereza Miranda,  
que são meus primeiros e eternos orientadores...

**Aline Miranda**

## AGRADECIMENTOS

A Deus, por estar sempre ao meu lado, sendo minha alegria, minha força e minha esperança...  
Aos meus pais, Vilson e Tereza, meus grandes exemplos, por doarem, inteiramente, suas vidas para minha felicidade e por serem sempre o motivo do meu sucesso. Vocês personificam a razão pela qual vale a pena lutar sempre!

A minha irmã Camila (*little*) por dividir comigo as alegrias e tristezas do percurso, por me ensinar a nunca desistir, e por me estimular a ser uma pessoa melhor a cada dia.

Ao meu noivo e anjo Antonio Neto, pelo companheirismo, incentivo, dedicação e amor!

A Daniele Santana (*Comp*) por ser minha amiga e companheira e, portanto, me fazer rir da vida sempre...

Ao Dr. Bernardo Galvão-Castro, coordenador do LASP, que sempre me incentivou e cultivou credibilidade e reconhecimento ao meu trabalho, além de todas as oportunidades concedidas.

Ao meu orientador, professor Dr. Luiz Alcântara, por todas as oportunidades, pela confiança integral, por tudo que me ensinou e pelos empurrões no abismo... aprendi muito com eles. Minha gratidão!

A todos os colegas e amigos do LASP pela constante presença, pelo total incentivo e pela confiança depositada. Em especial, a Ramon Moreau por todo suporte no núcleo de Bioinformática, a Andréa Gusmão, pelo apoio e pelos conselhos e a Viviana Olavarria, pela atenção e disponibilidade.

A Giselle Calasans (*Gisa*), Luciane Amorim (*Lu*) e Taisa Machado (*Pertubona*), pela amizade verdadeira, pela ajuda incondicional, por estar sempre ao meu lado, fazendo toda diferença, no percurso até aqui! A Filipe por todo profissionalismo, companheirismo e amizade!

As sempre colegas de pós-graduação Marcela Gomes (*Sheila*) e Caroline Urpia (*Carols*), por dividir comigo as alegrias e dificuldades de vencer esse desafio, além das máquinas de PCR, é claro.

A Amália Duarte (*Mamá*), Thais Bonfim (*Pertubona 2*), e Adriano Araújo (*Dri*), por completarem,  
com muito humor e carisma, a grande família que é o LASP.

A Dona Eugenia, Dona Beth, Cláudio e Rodrigo pelo suporte essencial durante todo o percurso.

A Catarina Bomfim pelo suporte na caracterização das seqüências depositadas no GenBank.

A Túlio de Oliveira por toda capacitação e apoio na utilização das ferramentas de Bioinformática.

A Pós-graduação em Biotecnologia em Saúde e Medicina Investigativa, especialmente à Taise  
Caíres pela competência, colocada a serviço, e atenção constante.

Ao Centro de Pesquisas Gonçalo Moniz, por toda estrutura e pelo apoio financeiro para o  
desenvolvimento dos trabalhos.

A Fundação Hemocentro de Rio Branco (Acre), por toda disponibilidade em facilitar o andamento  
das atividades e em especial a Juarez Pereira e a Noilson Lázaro pela realização da coleta das  
amostras.

A FAPESB pelo auxílio financeiro e pelo investimento nas atividades científicas desenvolvidas  
neste projeto.

A todos que contribuíram direta ou indiretamente para a realização deste trabalho.

## RESUMO

Em 1980, a partir de um paciente com linfoma cutâneo de células T, foi isolado o vírus linfotrópico de células T humanas tipo 1, o HTLV-1, que é conhecido, principalmente, por ser o agente etiológico de uma síndrome neurológica denominada TSP/HAM. No Brasil, estima-se que 2,5 milhões de indivíduos estejam infectados pelo HTLV-1, sendo a maior prevalência encontrada na população geral na cidade de Salvador, Bahia. Este estudo foi dividido em duas etapas: a primeira etapa refere-se a um estudo de caracterização do envelope viral utilizando ferramentas de bioinformática; enquanto que a segunda etapa caracteriza-se por ser um estudo de soroprevalência e análise dos novos isolados virais. O estudo de caracterização do envelope viral foi realizado para todas as seqüências nucleotídicas do gene *env* (n=15), já publicadas no Banco Mundial de seqüências, dando suporte ao programa brasileiro de DST-AIDS. Este estudo de caracterização se baseou na identificação de possíveis epítomos lineares, na caracterização físico-química, na identificação de mutações selecionadas positivamente, e por fim na procura por assinaturas correspondentes às modificações pós-traducionais. Inicialmente, realizamos a triagem de peptídeos previamente identificados no envelope viral, e em seguida a predição de epítomos nestas seqüências peptídicas. Em 12 peptídeos previamente publicados, foram encontrados 12 possíveis epítomos, cuja variação total na composição de aminoácido foi de 9% e 17% para os alelos de HLA classes I e II, respectivamente. Em 5 dos 12 epítomos a análise físico-química revelou que as mutações presentes foram capazes de aumentar o perfil de antigenicidade da região protéica que continha a mutação. A análise de domínios potenciais mostrou a perda de um sítio de fosforilação (PKC) no epítomo que contém a mutação D197N, e a perda de um sítio de N-glicosilação causada pelas mutações S246Y e V247I em outro epítomo. Além disso, a análise de pressão seletiva revelou 8 sítios selecionados positivamente ( $\omega = 9,59$ ), usando modelos de máxima verossimilhança. Este estudo ressalta a importância do envelope viral para o sucesso da infecção e para o desenvolvimento de protótipos vacinais, principalmente, porque foi possível identificar sítios selecionados positivamente, e epítomos com características conservativas e potenciais ligantes para um grande número de alelos de HLA. A segunda etapa do trabalho consistiu na caracterização genotípica de novos isolados do HTLV-1 na cidade de Rio Branco-Acre. Para contribuir com dados de soroprevalência do HTLV-1 no Brasil, e para discutir sobre a epidemiologia molecular do vírus, 219 doadores de sangue foram triados para a presença de anticorpos específicos contra o vírus. Um único caso de infecção (soroprevalência de 0,46%) foi detectado entre os doadores de sangue atendidos, durante o mês de Julho de 2004, na Fundação Hemocentro de Rio Branco-Acre. Seqüenciamos, então, a região LTR total referente a esse isolado viral (AC181), e submetemos à subtipagem, juntamente com outras 4 seqüências LTR (AC042, AC174, AC069, AC129), também geradas neste trabalho, provenientes, entretanto, de casos de infecção identificados por um trabalho anterior de soroprevalência nesta mesma população. Para os fragmentos totais da região LTR (AC181 e AC042), análises filogenéticas foram realizadas utilizando o programa PAUP\*, enquanto que os fragmentos parciais (AC174, AC069, AC129) foram genotipados através de uma ferramenta *online* de subtipagem. Três seqüências referentes ao gene *env* (ENVAC57, ENVAC69 e ENVAC204), também foram geradas, neste estudo, utilizando amostras HTLV-1 positivas provenientes de um estudo prévio de caracterização sorológica. Estas foram analisadas para a presença de modificações pós-traducionais. Análises filogenéticas demonstraram que tanto os isolados LTR total, como os parciais foram classificados como pertencentes ao subgrupo Transcontinental do subtipo Cosmopolita, dentro do grupo monofilético A da América Latina para os isolados AC181 e AC042. Comparando a diversidade genética das seqüências de HTLV-1, geradas neste trabalho, com outras seqüências virais em infecções em Salvador, Fortaleza e Peru, foi possível encontrar uma menor distância genética entre as cepas de Rio Branco, quando comparadas com as cepas de Fortaleza e do Peru, e

uma maior distância genética, quando comparadas com cepas de Salvador. A análise de domínios potenciais mostrou a presença de sítios de fosforilação para PKC nas posições 310-312aa e 342-344aa; um sítio de fosforilação para CK2 na posição 194-197aa; sítios de N-glicosilação nas posições 222-225aa, 244-247aa e 272-275aa; e por fim, um sítio de miristilação na posição 327-338aa. Os achados moleculares nos permitem sugerir uma possível origem Pré-Colombiana do vírus nesta população, sem, no entanto, descartar completamente a possibilidade de introdução Pós-Colombiana.



## ABSTRACT

The Human T cell lymphotropic virus type 1, HTLV-1, was isolated, in 1980, from a patient affected by a T cell cutaneous lymphoma, and it is mainly associated to the development of a neurological disorder called TSP/HAM. About 2.5 million of individuals are HTLV-1 infected, in Brazil, and the greater prevalence into general population is registered in Salvador city, Bahia state. This study was performed in two stages: the first one refers to an study of viral envelope characterization, using bioinformatics tools; while the second one, is a seroprevalence investigation and viral isolates characterization. The *env* gene characterization study was carried out to evaluate the molecular pattern of all available Brazilian human T-cell lymphotropic virus type 1 Env (n = 15) nucleotide sequences via epitope prediction, physico-chemical analysis, and protein potential sites identification, giving support to the Brazilian AIDS vaccine program. In 12 previously described peptides of the Env sequences, 12 possible epitopes were found. The total variation on the amino acid composition was 9% and 17% for human leukocyte antigen (HLA) class I and class II Env epitopes, respectively. In 5 of the 12 Env epitopes the physico-chemical analysis demonstrated that the mutations magnified the antigenicity profile into the mutation region. The potential protein domain analysis of Env sequences showed the loss of a CK-2 phosphorylation site caused by D197N mutation in one epitope, and the loss of a N-glycosylation site caused by S246Y and V247I mutations in another epitope. Besides, the analysis of selective pressure have found 8 positive selected sites ( $\omega = 9.59$ ) using the codon-based substitution models and maximum-likelihood methods. These studies underscore the importance of this Env region for the virus fitness, for the host immune response and, therefore, for the development of vaccine candidates. To contribute further with the HTLV-1 seroprevalence in Brazil, and to discuss the virus molecular epidemiology in this Amazon population (Rio Branco-Acre), 219 blood donors were screened for HTLV-1-specific antibodies. It was only detected a single case of infection (0.46% seroprevalence) among blood donors, screened during July 2004, at FUNDACRE. We have submitted this unique HTLV-1 positive sample (AC181) and four others positive samples (AC042, AC174, AC069, AC129) originated in a previous serological study, in the same population, to the sequencing of the complete LTR region, and submitted to genotyping. To assess molecular epidemiology, Neighbor-joining and Maximum Likelihood phylogenetic analyses of complete LTR region sequences (AC181 and AC042) were performed with PAUP\* software, while the partial LTR fragments (AC129, AC174 and AC069) were genotyped using the online subtyping tool. Three envelope sequences (ENV57, ENV69, ENV204) were also generated, in this study, from HTLV-1 positive samples from a previous serological study, and analyzed using the Prosite tool to determinate potential protein sites. Phylogenetic analysis demonstrated that the total and partial LTR strains belong to the Transcontinental subgroup of the Cosmopolitan subtype, inside the Latin American cluster A, for the isolates AC181 and AC042. Calculating the genetic diversity among the HTLV-1 sequences, generated in this report, and sequences described in infection cases in Salvador, Fortaleza e Peru, it was possible to find a close relationship between Rio Branco and Fortaleza or Peru sequences. The potential protein site analysis showed, that all env sequences encoded two PKC phosphorylation sites at amino acid (aa) positions 310-312 and 342-344, one CK2 phosphorylation site at 194-197aa, three N-glycosylation sites at 222-225aa, 244-247aa and 272-275aa, and a single N-myristylation site at 327-338aa. These findings allow to infer about a possible virus introduction in this population through a Pre-Columbian human migration, although we can not exclude a possible Post-Columbian introduction.

## LISTA DE FIGURAS

<b>Figura 1</b>	<b>Estrutura Morfológica do HTLV: desenho esquemático</b> (Adaptado de SALEMI, 1999) .....	<b>15</b>
<b>Figura 2</b>	<b>Organização Genômica do HTLV-1</b> (MATSUOKA & JEANG, 2007).....	<b>17</b>
<b>Figura 3</b>	<b>Desenho esquemático do heterodímero formado pelas glicoproteínas SU e TM</b> (Adaptado de MANEL <i>et al.</i> , 2004).....	<b>19</b>
<b>Figura 4</b>	<b>Esquema ilustrativo da interação do heterodímero gp46-gp21 na fusão celular.</b> (Adaptado: <a href="http://www.dundee.ac.uk/biomedres/brighty.htm">www.dundee.ac.uk/biomedres/brighty.htm</a> ).....	<b>22</b>
<b>Figura 5</b>	<b>Ciclo de Replicação do HTLV</b> (SALEMI, 1999) .....	<b>23</b>
<b>Figura 6</b>	<b>Mapa da prevalência do HTLV-1 no Brasil e no mundo</b> (Adaptado de PROIETTI <i>et al.</i> , 2005 e CATALAN-SOARES <i>et al.</i> , 2005) .....	<b>27</b>
<b>Figura 7</b>	<b>Representação esquemática dos diferentes resultados no <i>Western Blot</i>....</b>	<b>35</b>
<b>Figura 8</b>	<b>Delineamento do estudo de caracterização molecular dos genes <i>env</i> e <i>pol</i> em seqüências publicadas no GenBank.....</b>	<b>40</b>
<b>Figura 9</b>	<b>Delineamento do estudo genotípico e epidemiológico de isolados do HTLV-1 de Rio Branco-Acre .....</b>	<b>44</b>

## LISTA DE ABREVIATURAS E SIGLAS

3' .....	Região carboxi-terminal do ácido nucléico
5' .....	Região amino-terminal do ácido nucléico
aa .....	Aminoácido
Anti-IgG.....	Anticorpo do isotipo IgG
ANVISA.....	Agência Nacional de Vigilância Sanitária
Asp-His-Ile-Leu- Glu-Pro-Ser-Ile-Pro	Asparagina-Histidina-Isoleucina-Leucina-Glutamato-Prolina-Serina-Isoleucina-Prolina
ATK1 .....	Cepa referência do genoma do HTLV-1
ATL .....	Leucemia/linfoma de células T do adulto ( <i>Adult T cell Leukemia</i> )
CA.....	Capsídeo
cAMP .....	Mensageiro secundário na transdução de sinal
CD4+ .....	Linfócitos T CD4+
CD45RO .....	Linfócitos T de memória
CD8+ .....	Linfócitos T CD8+
CREB/ATF .....	Fator de Transcrição
CTL .....	Linfócito T citotóxico ( <i>Citotoxic T Lymphocyte</i> )
DNA .....	Ácido desoxirribonucléico
EIA .....	Ensaio Imunoenzimático ( <i>Enzyme Imune Assay</i> )
ELISA.....	<i>Enzyme linked Imuno Sorbent Assay</i>
env .....	Envelope
gag .....	Grupo antigênico
Gln-Glu-Gln-Cys- Arg-Phe-Pro-Asn- Ile-Thr .....	Glicina-Glutamato-Glicina-Cisteína-Arginina-Phenilalanina-Prolina-Asparagina-Isoleucina-Threonina
GLUT-1.....	Molécula transportadora de glicose
gp21 .....	Glicoproteína transmembrana
gp46 .....	Glicoproteína de superfície
HBZ.....	Gene codificado pela fita complementar negativa
HIF-1 .....	Fator Indutor de Hipóxia do tipo 1 ( <i>Hypoxia-Inducible Factor 1</i> )
HIV .....	Vírus da Imunodeficiência Humana ( <i>Human Immunodeficiency Virus</i> )
HLA .....	Antígeno Leucocitário humano ( <i>Human Leucocitary Antigen</i> )
HTLV-1.....	Vírus Linfotrópico de células T humanas tipo 1 ( <i>Human T cell Lymphotropic vírus type 1</i> )
HTLV-2 .....	Vírus Linfotrópico de células T humanas tipo 2 ( <i>Human T cell Lymphotropic vírus type 2</i> )
HTLV-3.....	Vírus Linfotrópico de células T humanas tipo 3 ( <i>Human T cell Lymphotropic vírus type 3</i> )
HTLV-4 .....	Vírus Linfotrópico de células T humanas tipo 4 ( <i>Human T cell Lymphotropic vírus type 4</i> )
IBGE.....	Instituto Brasileiro de Geografia e Estatística
IL-2 .....	Interleucina do tipo 2
IN .....	Integrase
K55 .....	Peptídeo da gp46 específico do HTLV-2
kb .....	Kilobases
kDa .....	Kilo Daltons
LASP .....	Laboratório Avançado de Saúde Pública

LRT.....	Razão de Verossimilhança ( <i>Likelihood Ratio Test</i> )
LTR.....	Extremidades em repetições longas ( <i>Long Terminal Repeat</i> )
MA .....	Proteína da matriz
ML .....	Máxima Verossimilhança ( <i>Maximum Likelihood</i> )
mRNA .....	Molécula de ácido ribonucléico mensageiro
MTA-1 .....	Peptídeo da gp46 específico do HTLV-1
MTOC.....	Organização Microtubular ( <i>Microtubule-organizing Center</i> )
NF-κB .....	Fator de Transcrição ( <i>Nuclear Factor κB</i> )
NJ .....	Agrupamento de vizinhos ( <i>Neighbor Joining</i> )
OMS .....	Organização Mundial de Saúde
ORF.....	Fase de Leitura aberta ( <i>Open Reading Frame</i> )
P12, p30 e p13.....	Proteínas acessórias
p15 .....	Proteína do nucleocapsídeo
p19 .....	Proteína da matriz
p24 .....	Proteína do capsídeo
pb .....	Pares de bases
PCR .....	Reação em Cadeia da Polimerase ( <i>Polymerase Chain Reaction</i> )
pol .....	Polimerase
Pro .....	Protease
pX .....	Gene regulatório
RBD .....	Domínio de ligação ao receptor ( <i>Receptor Binding Domain</i> )
RNA.....	Ácido Ribonucléico
RNAseH .....	Enzima específica para digestão de RNA
SEMTA .....	Serviço Especial de Mobilização de Trabalhadores para a Amazônia
SEMTA.....	Serviço Especial de Mobilização de Trabalhadores para Amazônia
SRF .....	Fator de Transcrição ( <i>Serum Response Factor</i> )
STLV .....	Vírus linfotrópico de células T em Símios ( <i>Simian T cell Lymphotropic virus</i> )
STLV-1 .....	Vírus linfotrópico de células T em Símios tipo 1 ( <i>Simian T cell Lymphotropic virus type 1</i> )
SU.....	Proteína de superfície
TATA box .....	Região que antecede o sítio de transcrição
TAX.....	Proteína viral regulatória
TCLE .....	Termo de Consentimento Livre e Esclarecido
TM .....	Proteína transmembrana
TR .....	Transcriptase reversa
TSP/HAM.....	Paraparesia Espástica Tropical ( <i>Tropical Spastic Paraparesis</i> )/ Mielopatia Associada ao HTLV ( <i>HTLV Associated Myelopathy</i> )
WB .....	Teste sorológico confirmatório para infecção pelo HTLV-1 ( <i>Western Blot</i> )

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>14</b>
<b>1.1 Vírus Linfotrópico de Células T Humanas do Tipo 1 (HTLV-1)</b>	<b>14</b>
<b>1.2 Estrutura da Partícula Viral</b>	<b>15</b>
<b>1.3 Estrutura Genômica do HTLV</b>	<b>16</b>
<b>1.4 Ciclo de Replicação Viral</b>	<b>21</b>
<b>1.5 Epidemiologia da Infecção</b>	<b>24</b>
<b>1.6 Doenças Associadas ao HTLV-1</b>	<b>29</b>
<b>1.7 Resposta Imune na Infecção pelo HTLV-1</b>	<b>31</b>
<b>1.8 Diagnóstico Sorológico</b>	<b>33</b>
<b>2. JUSTIFICATIVA .....</b>	<b>36</b>
<b>3. OBJETIVOS .....</b>	<b>38</b>
<b>3.1 Objetivo Geral</b>	<b>38</b>
<b>3.2 Objetivos Específicos</b>	<b>38</b>
<b>4. METODOLOGIA .....</b>	<b>40</b>
<b>4.1 Caracterização molecular de seqüências dos genes <i>env</i> e <i>pol</i>, disponíveis no Banco Mundial de seqüências</b>	<b>40</b>
4.1.1- Delineamento do estudo	40
4.1.2- Caracterização dos objetos de estudo	41
4.1.3- Identificação de possíveis epítomos lineares	41
4.1.4- Caracterização físico-química	42
4.1.5- Identificação de domínios potenciais	42
4.1.6- Identificação das mutações	42
4.1.7- Cálculo de pressão seletiva	43
<b>4.2 Estudo genotípico e epidemiológico de isolados do HTLV-1 de Rio Branco-Acre</b>	<b>43</b>
4.2.1- Delineamento do estudo	43
4.2.2- Caracterização da população	44
4.2.3- Sorologia para HTLV-1/2	45
4.2.4- Extração de DNA	45
4.2.5- Reação em cadeia da polimerase (PCR)	45
4.2.6- Seqüenciamento	46
4.2.7- Análise filogenética	46
4.2.8- Cálculo da diversidade genética	47
4.2.9- Identificação de domínios potenciais	47
4.2.10- Aspectos éticos	48

<b>5.</b>	<b>RESULTADOS .....</b>	<b>49</b>
<b>5.1</b>	<b>Caracterização molecular de seqüências dos genes <i>env</i> e <i>pol</i> previamente publicadas</b>	<b>49</b>
5.1.1-	Artigo Publicado: Mota-Miranda, AC; De Oliveira, T; Moreau, DR; Bonfim, C; Galvão-Castro, B; Alcântara, LCJ. <i>Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design. Mem Inst Oswaldo Cruz</i> 102(6): 741-749, 2007	50
<b>5.2</b>	<b>Caracterização genotípica de novos isolados do HTLV-1 de Rio Branco</b>	<b>72</b>
5.2.1-	Artigo Publicado: Mota-Miranda, AC; Araújo, SP; Dias, JP; Colin, DD; Kashima, S; Covas, DT; Tavares-Neto, J; Galvão-Castro, B; Alcântara, LCJ. <i>HTLV-1 infection in blood donors from Western Brazilian Amazon region: seroprevalence and molecular study of viral isolates. J Med Virol</i> 80(11):1966-1971, 2008	73
<b>6.</b>	<b>DISCUSSÃO .....</b>	<b>92</b>
<b>7.</b>	<b>CONCLUSÃO .....</b>	<b>95</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>97</b>

## 1. INTRODUÇÃO

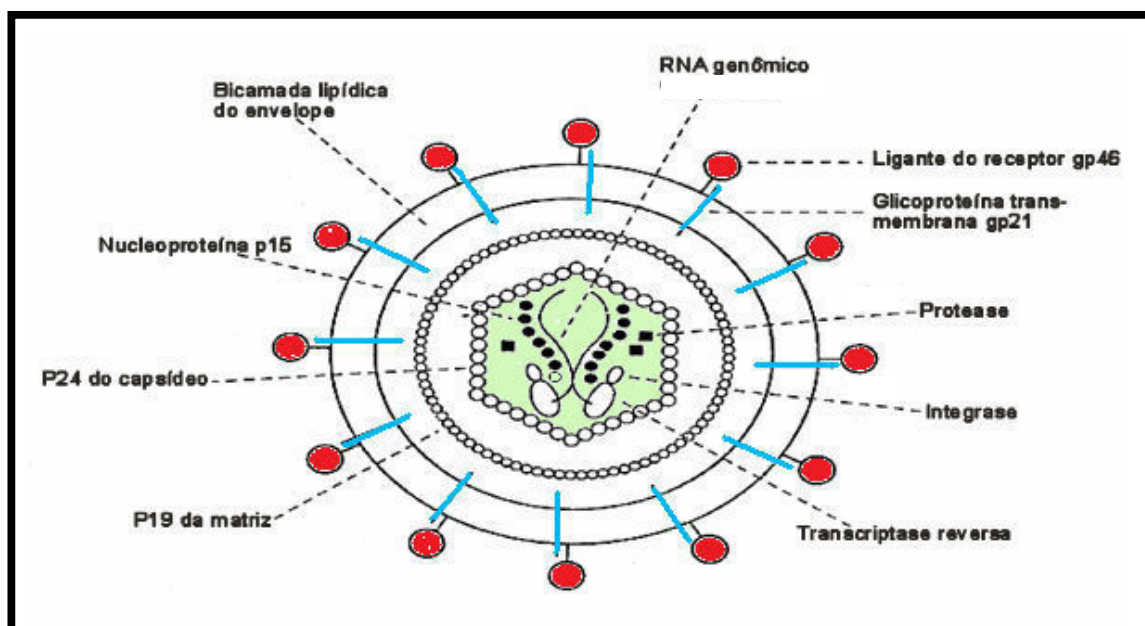
### 1.1 Vírus Linfotrópico de Células T Humanas do Tipo 1 (HTLV-1)

Em 1980, a partir de um paciente com linfoma cutâneo de células T, foi isolado o vírus linfotrópico de células T humanas tipo 1 (*Human T-cell Lymphotropic Virus Type 1*-HTLV-1) (POIESZ *et al.*, 1980). O HTLV-2 (*Human T-cell Lymphotropic Virus Type 2*), por sua vez, foi identificado pela primeira vez em 1982, numa linhagem de células T imortalizadas de um paciente com tricoleucemia (KALYANARAMAN *et al.*, 1982). Mais recentemente, isolados virais descritos como HTLV-3 (*Human T-cell Lymphotropic Virus Type 3*) e HTLV-4 (*Human T-cell Lymphotropic Virus Type 4*) foram isolados pela primeira vez em indivíduos da África Central (WOLFE *et al.*, 2005).

Inicialmente, a infecção pelo HTLV-1 foi associada ao desenvolvimento da leucemia/ linfoma de células T de adulto (*Adult T-cell leukemia/lymphoma*-ATL) (POIESZ *et al.*, 1980), cujas primeiras ocorrências foram relatadas em indivíduos infectados no Japão. Entre os anos de 1985 e 1986, foram descritas duas outras manifestações clínicas associadas ao HTLV-1: paraparesia espástica tropical (*Tropical Spastic Paraparesis*-TSP) e mielopatia associada ao HTLV (*HTLV Associated Myelopathy*-HAM), em infecções na Martinica e no Japão, respectivamente (GESSAIN *et al.*, 1985; OSAME *et al.*, 1986). No entanto, alguns anos depois, foi possível concluir tratar-se da mesma etiologia, de forma que, atualmente, o HTLV-1 é conhecido como o agente etiológico de uma síndrome neurológica denominada TSP/HAM. Adicionalmente, artropatias (NISHIOKA, 1996), polimiosites (MORGAN *et al.*, 1989), uveítes (MOCHIZUKI *et al.*, 1996), e dermatites infectivas (LA GRENADÉ, 1996) também têm sido relatadas como sendo manifestações associadas à infecção viral.

## 1.2 Estrutura da Partícula Viral

A estrutura da partícula viral é constituída, basicamente, por um envelope, uma matriz protéica e um nucleocapsídeo. A morfologia do vírus é de esférica a pleomórfica, estrutura similar a de outros retrovírus. O envelope constitui o único complexo protéico presente na superfície do vírion, consistindo de duas subunidades protéicas glicosiladas: a proteína de superfície (SU), denominada gp46, e a proteína transmembrana (TM), denominada gp21, que ancora a gp46 (**Fig. 1**). Logo após a membrana do envelope encontra-se a matriz viral que é composta pela proteína da matriz (MA), ou p19. O capsídeo (CA), por sua vez, apresenta simetria icosaédrica, constitui o cerne da partícula viral e é composto principalmente pela proteína p24.



**Fig. 1-Estrutura Morfológica do HTLV: desenho esquemático** (Adaptado de SALEMI, 1999).



O genoma viral, localizado no interior do capsídeo, é composto por duas fitas de RNA, 8-9 kilobases (kb), as quais estão associadas a várias pequenas proteínas, chamadas de proteínas do nucleocapsídeo (p15). Ainda associadas ao genoma viral estão proteínas importantes no processo de integração do DNA proviral ao genoma da célula hospedeira (transcriptase reversa-TR e integrase-IN) e em eventos catalíticos que precedem o ciclo de replicação viral (protease e RNaseH).

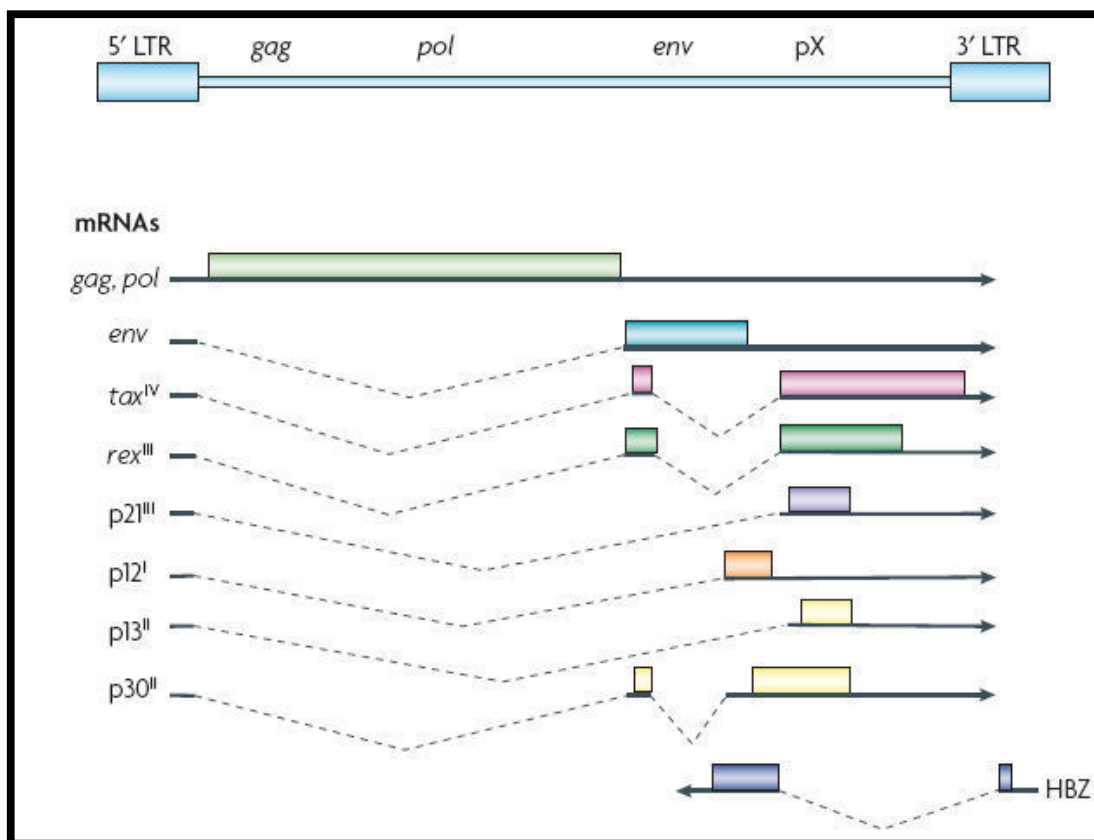
### 1.3 Estrutura Genômica do HTLV

Com uma organização similar a de outros retrovírus, o genoma do HTLV possui os genes *gag* (grupo antigênico), *pol* (polimerase) e *env* (envelope) além de uma seqüência próxima à extremidade 3' conhecida como região pX, a qual contém, entre outros, os genes reguladores *tax* e *rex*, responsáveis pela transativação e regulação da expressão gênica (**Fig. 2**). Completando a estrutura do genoma, estão as regiões flanqueadoras LTR (*Long Terminal Repeat*) que exercem função essencial na incorporação do DNA proviral ao genoma da célula hospedeira, bem como na regulação da transcrição de genes virais (GREEN & CHEN, 2001). As regiões flanqueadoras estão presentes nas duas extremidades (5' e 3') do DNA proviral e são constituídas por três subunidades: U3, R e U5.

Existem pelo menos três elementos essenciais na região U3: 1-três repetições de 21 pares de bases, necessárias para a transativação mediada por TAX (BRADY *et al.*, 1987); 2- o TATA box, que antecede o sítio de iniciação transcricional; 3- e o sinal de poliadenilação. Além disso, essas regiões flanqueadoras são altamente variáveis e são, portanto, utilizadas como fonte de informação em estudos de epidemiologia molecular.

No processo de transcrição, três moléculas de mRNA são produzidas: 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 LTR3'; o mRNA sub-genômico, sintetizado a partir de uma única etapa de processamento e codificante do produto do gene *env*; e um segundo mRNA sub-genômico, 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*).



**Fig. 2 – Organização genômica do HTLV-1 (MATSUOKA & JEANG, 2007).** Indicação dos genes e principais moléculas de mRNA produzidas durante a transcrição.

O gene *gag* está compreendido entre os nucleotídeos 824 e 2113 (1218pb) no genoma viral do protótipo ATK1 (Número de Acesso J02029) (SEIKI *et al.*, 1983), e seu final 3' sobrepõe-se ao início da ORF que codifica para protease. Essa região é inicialmente traduzida como um precursor

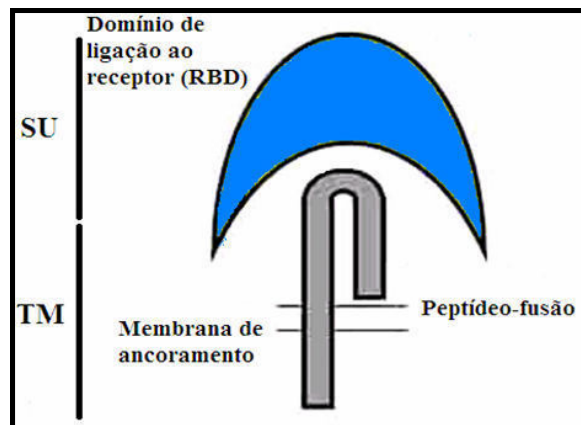
poliprotéico, cuja subsequente clivagem dá origem às proteínas estruturais maduras do gene *gag*: a proteína da matriz de 19kDa (p19), a proteína do capsídeo de 24 kDa (p24) e a proteína do nucleoproteína de 15kDa (p15).

A protease (PRO), por sua vez, é codificada por desvio de leitura dos genes *gag* e *pol*, entre os nucleotídeos 1849 e 2757 (909pb) no genoma viral de referência (Número de acesso D13784) (MALIK *et al.*, 1988). A protease atua sobre as cadeias poliprotéicas, clivando-as para formação das proteínas estruturais maduras encontradas na partícula viral.

O gene *pol* compreende os nucleotídeos 2520 a 5207 (2688pb) do genoma referência ATK1 do HTLV-1, sendo que a extremidade 5' do gene *pol* codifica a transcriptase reversa e a sequência *downstream* codifica a integrase. A TR é a enzima responsável pela síntese do DNA viral a partir do RNA de fita simples e, portanto, tem papel fundamental na fase inicial do ciclo de multiplicação do retrovírus.

O gene *env* que corresponde à sequência de nucleotídeos 5203 a 6669 (1467pb) do genoma viral referência ATK1 (Número de Acesso J02029) (SEIKI *et al.*, 1983), codifica a proteína precursora do envelope viral (ENV). A proteína precursora ENV é clivada, entre os nucleotídeos 6117 e 6118, para gerar os produtos maduros, a glicoproteína de superfície de 46kDa (gp46-SU) e a proteína transmembrana de 21kDa (gp21-TM).

Como no caso de outros retrovírus, ocorre associação não covalente entre a proteína de superfície e a proteína transmembrana, de modo que esta última ancora a proteína de superfície no envelope da partícula viral (**Fig. 3**) (DELAMARRE *et al.*, 1996; MANEL *et al.*, 2004).



**Fig. 3- Desenho esquemático do heterodímero formado pelas glicoproteínas SU e TM.** (Adaptado de MANEL *et al.*, 2004).

As glicoproteínas gp46 e gp21 estão diretamente associadas ao reconhecimento celular, e à conseqüente entrada do vírus na célula, portanto, inúmeros estudos têm se dedicado a investigar quais seriam as regiões gênicas envolvidas no exercício dessa função. Sagara e col. (1996) identificaram peptídeos sintéticos, que contendo seqüências específicas dos aminoácidos 197 a 205 (Asp-His-Ile-Leu-Glu-Pro-Ser-Ile-Pro) em gp46 e 397 a 406 (Gln-Glu-Gln-Cys-Arg-Phe-Pro-Asn-Ile-Thr) em gp21, foram capazes de inibir a formação de sincício induzido pela infecção pelo HTLV-1, sugerindo se tratar de dois domínios funcionais envolvidos na função do envelope viral.

As glicoproteínas do envelope são altamente conservadas entre os diferentes isolados do HTLV-1, no entanto, substituições nucleotídicas na região gênica que codifica para estas proteínas, podem influenciar na infectividade viral, no tropismo celular, na taxa de replicação e na latência ou resposta aos mecanismos efetores da resposta imunológica, como produção de anticorpos (SZUREK *et al.*, 1988; PAQUETTE *et al.*, 1989).

Rosenbeg e col. (1998), por sua vez, ao estudar características conservativas das glicoproteínas do envelope, identificaram outros dois domínios imunodominantes em gp46: o primeiro se situa entre

os aminoácidos 75 e 101 e o segundo, na região mais central da proteína, compreendendo os aminoácidos 181 a 208. Desta forma, uma única troca de aminoácido no resíduo 197 (Aspartato) reduz drasticamente a fusão celular, de forma semelhante, mutações localizadas na região N-terminal (em torno do aminoácido 90) da gp46 podem dificultar diretamente a formação de sincício e a entrada do vírus na célula hospedeira. Além disso, no domínio extracelular da gp21 também foi possível caracterizar duas regiões funcionalmente distintas localizadas na porção N-terminal que inclui motivos estruturais, *leucine zipper-like*, que se relacionam com a fusão entre membranas e com a transmissão célula-célula.

As proteínas regulatórias Tax e Rex, codificadas pela transcrição do gene *pX*, têm função importante na dinâmica viral, de modo que cabe à proteína Tax a ativação de genes celulares que codificam fatores de transcrição, como NF- $\kappa$ B, fatores ligantes dos elementos responsivos ao cAMP, IL-2 e receptor de IL-2. Além disso, Tax abriga epítomos (11-19 aa) importantes que fazem dela um alvo antigênico dominante no reconhecimento por linfócitos T citotóxicos na maioria dos indivíduos. A proteína Rex, por sua vez, responde pela regulação dos níveis de expressão das proteínas virais, funcionando, portanto, como um modulador da transcrição da fase inicial até a fase produtiva no ciclo de replicação viral.

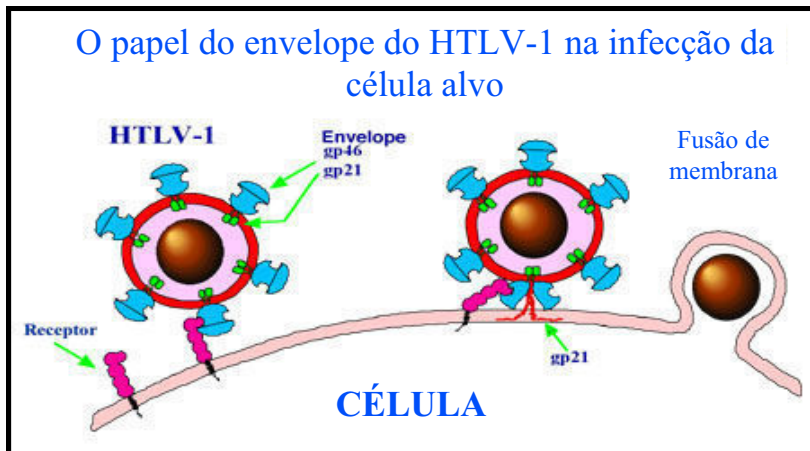
O gene HBZ, por sua vez, é codificado pela fita complementar negativa (3'  $\rightarrow$  5') do HTLV-1, e contém um domínio *leucine zipper*. O HBZ interage diretamente com c-Jun ou Jun-B, ou maximiza sua degradação, resultando na supressão da transcrição viral mediada por tax, iniciada a partir da região LTR (MATSUOKA, 2005). Desta forma, experimentos conduzidos em modelos animais, utilizando clones mutantes do HTLV-1 forneceram evidências de que os genes acessórios p12, p30, rex, p13 e

HBZ contribuem para a persistência da infecção *in vivo*, através do efeito direto dos produtos gênicos na replicação viral e na proliferação de células infectadas (MATSUOKA & JEANG, 2007).

A alta estabilidade genotípica do HTLV-1 é provavelmente devido à replicação viral por meio da expansão clonal das células infectadas, e ao uso mínimo da transcriptase reversa (WATTEL *et al.*, 1995; BANGHAM & OSAME, 2005). O genoma do HTLV-1 por ser estável, é amplamente utilizado como bom marcador molecular para traçar eventos de migração das populações humanas ancestrais, bem como para o entendimento dos mecanismos da evolução viral e monitoramento dos meios de transmissão (GESSAIN *et al.*, 1992).

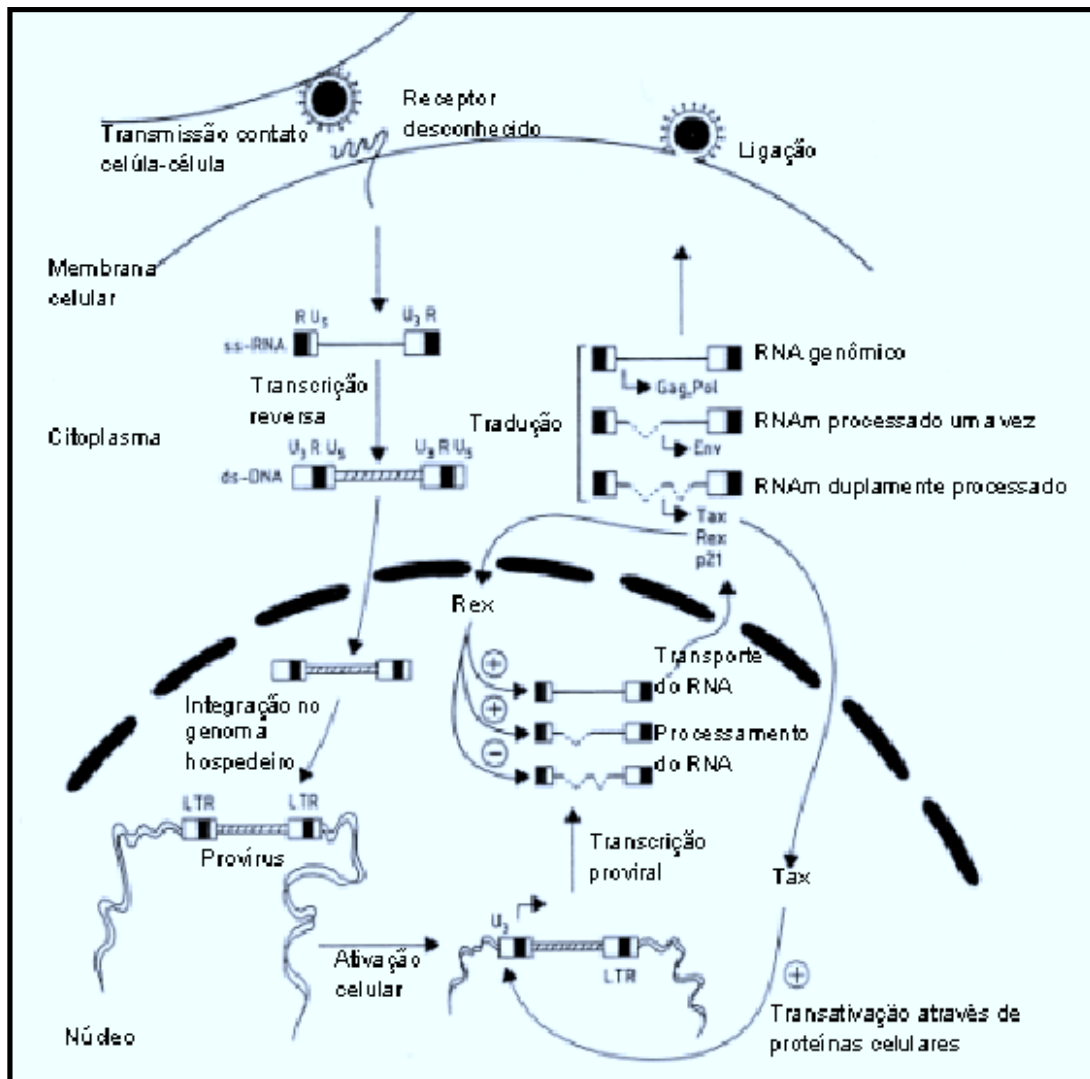
#### **1.4 Ciclo de Replicação Viral**

A primeira etapa do ciclo de replicação é a ligação do vírus na célula hospedeira, entre o domínio de ligação do receptor (*Receptor Binding Domain-RBD*), na glicoproteína de superfície do vírus, e a molécula de GLUT-1, que é um transportador de glicose, na superfície da membrana celular (BATTINI *et al.*, 2003; MANEL *et al.*, 2004; COSKUN & SUTTON, 2005). O reconhecimento celular provoca modificações conformacionais na glicoproteína transmembrana de modo que se efetuem a subsequente fusão entre membranas e a internalização do capsídeo viral no citoplasma da célula hospedeira (**Fig. 4**).



**Fig. 4-** Esquema ilustrativo da interação do heterodímero gp46-gp21 na fusão celular. (Adaptado: [www.dundee.ac.uk/biomedres/brighty.htm](http://www.dundee.ac.uk/biomedres/brighty.htm))

A transcrição do genoma viral de RNA para DNA, pela enzima TR ocorrerá dentro do cerne viral usando como oligonucleotídeo iniciador o tRNA<sup>Pro</sup>, que será, tardiamente, degradado pela RNaseH. Subsequentemente, o DNA viral entra no núcleo, e é inserido no genoma da célula hospedeira formando o provírus, função exercida pela integrase viral. O processo de integração do provírus marca o final da fase precoce do ciclo de multiplicação do vírus e inicia a fase tardia que é mediada por enzimas do hospedeiro (SEIKI *et al.*, 1984). Nesta última etapa, ocorre a síntese do RNA viral tendo como DNA molde o provírus integrado. A síntese do RNA viral leva à formação de um longo transcrito primário, que é processado para formar as moléculas de mRNA e RNA genômico, de forma que as proteínas virais são, então, traduzidas nos ribossomos celulares e maturadas (**Fig. 5**).



**Fig. 5-Ciclo de Replicação do HTLV (SALEMI, 1999).**

Em modelos clássicos de replicação dos retrovírus, as proteínas recém sintetizadas são montadas em novas partículas virais infectivas, que brotam da superfície celular para reiniciar o ciclo em outra célula hospedeira.



No entanto, ensaios funcionais disponíveis, até agora, sugerem que o HTLV-1 é pouco replicativo e que a replicação viral *in vivo* ocorre, principalmente, devido à expansão clonal das células infectadas, via mitose, do que via transcrição reversa (WATTEL *et al.*, 1995; CIMARELLI *et al.*, 1996). Além disso, o vírus induz eventos de polarização das células facilitando a passagem viral, num fenômeno conhecido como sinapse viral (BANGHAM, 2003). Quando uma célula infectada entra em contato com outra célula, porém, não infectada há a formação de um centro de organização microtubular (MTOC ou *Microtubule-organizing center*) que é polarizado na junção célula-célula, formando a interface onde ocorrerá a sinapse virológica. A formação desta estrutura permite o acúmulo de proteínas de gag e de material genômico (RNA), na interface de sinapse, culminado com a passagem desse material para a célula não infectada (MATSUOKA & JEANG, 2007). Manel e col. (2005) demonstraram, inclusive, que nestas áreas de “sinapse viral” há um acúmulo de moléculas de GLUT 1, conhecidas por seu papel no reconhecimento celular.

### **1.5 Epidemiologia da Infecção**

A transmissão do HTLV-1 ocorre, principalmente, por três vias: horizontal (contato sexual), sendo a ocorrência mais freqüente do homem para a mulher; vertical (da mãe para o filho), caracterizada, principalmente, pela amamentação, ou ainda por transmissão transplacentária, ou durante o parto; e parenteral, ocorrendo através da transfusão de sangue contaminado e seus produtos, bem como do uso de seringas, ou perfuro cortantes contaminados (CARNEIRO-PROIETTI *et al.*, 2002). A transmissão do HTLV-1 é menos eficiente que a do vírus da imunodeficiência humana (HIV), devido à baixa carga proviral e ao fato de a infecção ser dependente do contato célula/célula (BANGHAM, 2003).

No entanto, não se pode descartar a influência que a rota de transmissão pode ter no desenvolvimento da resposta imune antiviral, e conseqüentemente no desenvolvimento de doenças associadas, como é o caso da ATL que tem sido associada à transmissão vertical. Desenhos experimentais utilizando modelos murinos, relatam que a falta de resposta imune, avaliada através da produção de anticorpos, associada à intensa resposta linfo-proliferativa e à possível transformação de células infectadas, na mucosa oral, poderia explicar o desenvolvimento de ATL em indivíduos verticalmente infectados (YASUNAGA & MATSUOKA, 2007).

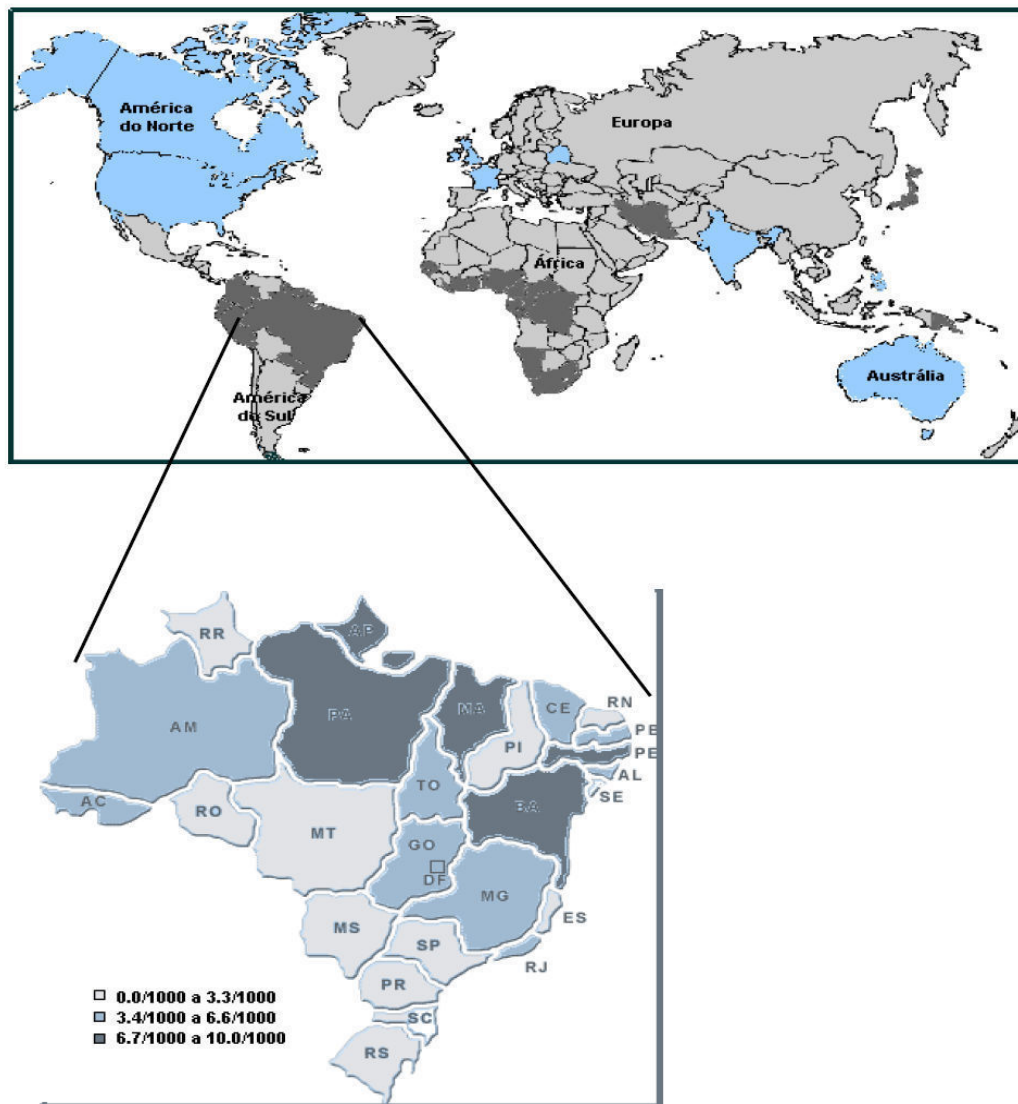
Em quase 30 anos desde o isolamento do vírus, as taxas de prevalência da infecção se relacionam, de forma crescente, com características geográficas, composição sócio-demográfica, e comportamentos individuais de risco (GALVÃO-CASTRO *et al.*, 1997).

Dados epidemiológicos mostram que a infecção pelo HTLV-1 tem distribuição mundial (DE THE & KAZANJI, 1996), no entanto, algumas regiões são consideradas áreas endêmicas: sudoeste do Japão (YAMAGUCHI, 1994; MUELLER *et al.*, 1996), países no Caribe, (HANCHARD *et al.*, 1990), na África sub-Saara (GESSAIN & DE THE, 1996) e áreas localizadas no Irã e Melanésia (MUELLER, 1991). Taxas de prevalência, mais baixas, são encontradas em países da América do Sul, como no Brasil onde se estima que 2,5 milhões de indivíduos estejam infectados pelo HTLV-1 (CARNEIRO-PROIETTI *et al.*, 2002; CATALAN-SOARES *et al.*, 2004).

No Brasil, apesar dos dados epidemiológicos serem relativamente escassos, já que, em sua maioria, são restritos quase que exclusivamente à descrição da prevalência em populações específicas, também se pode observar o fenômeno de distribuição heterogênea do vírus, sendo as maiores prevalências observadas nos estados do Maranhão (São Luis- 10.0/1000 doadores de sangue), Bahia (Salvador-1,8% na população geral, ou 9.4/1000 doadores de sangue), Pará (Belém-9.1/1000 doadores

de sangue) e Pernambuco (Recife-7.5/1000 doadores de sangue) (**Fig. 6**) (DOURADO *et al.*, 2003; CATALAN-SOARES *et al.*, 2005).

A infecção pelo HTLV-1 é endêmica e largamente distribuída entre grupos indígenas e afro-descendentes em países da América do Norte, América Central e América do Sul. No Brasil, diversos grupos indígenas estão distribuídos por oito estados, que ocupam toda extensão da região amazônica (ISHAK *et al.*, 2003). Estudos epidemiológicos evidenciam que a região amazônica é uma extensa área endêmica para a ocorrência do HTLV-2 (ISHAK *et al.*, 1995) e mais recentemente, foi possível identificar a presença de HTLV-1 em tribos Wayampí (SHINDO *et al.*, 2002). Colin e col. (2003), em estudo com doadores de sangue da região amazônica, Rio Branco-Acre, estimaram ser de 0,08% e 0,03% as taxas de prevalência para o HTLV-1 e HTLV-2, respectivamente. Estudos de epidemiologia molecular de cepas virais demonstram a estreita relação filogenética entre isolados de HTLV-1 da América Latina e de isolados da África do Sul, quando comparado com isolados do oeste da África, dados que refletem a intensa entrada de sul-africanos no Brasil, durante o tráfico de escravos, durante os séculos XIV a XIX, e a maior prevalência do vírus em população afro-descendente no Brasil (DOURADO *et al.*, 2003; MOTA *et al.*, 2007).



**Fig. 6-Mapa da prevalência do HTLV-1 no Brasil e no mundo** (Adaptado de PROIETTI *et al.*, 2005 e CATALAN-SOARES *et al.*, 2005).

A endemicidade do HTLV-1 e HTLV-2, em algumas populações vivendo em áreas remotas do globo, sugerem a possibilidade desses vírus terem infectado populações humanas desde milhares de anos atrás. Através da análise de DNA mitocondrial, estima-se que a separação entre as populações humanas africanas e não-africanas ocorreu a cerca de 75.000-287.000 anos (CAVALLI-SFORZA *et*

*al.*, 1994; REICH & GOLDSTEIN, 1998), e várias observações mostraram que este fluxo gênico ocorreu normalmente dos pigmeus para as populações vizinhas (CAVALLI-SFORZA *et al.*, 1994). Baseado nessas evidências pode-se inferir que as infecções pelo HTLV-1 e HTLV-2 entre os pigmeus são as mais antigas ou resultou de transmissões interespecies, mais recentes, a partir do STLV. Transmissões interespecies freqüentes do STLV-1 de símios para humanos já foram demonstradas na África (VANDAMME *et al.*, 1994; LIU *et al.*, 1996; MAHIEUX *et al.*, 1998; SALEMI *et al.*, 1998).

Estudos genéticos e suas contribuições indicam pelo menos duas hipóteses para dar suporte à origem do HTLV-1 na América. A primeira hipótese é a Pré-Colombiana, baseada na migração de populações oriundas do norte da Ásia, através do Estreito de Bering, há cerca de 15.000-35.000 anos atrás (LAIRMORE *et al.*, 1990; NEEL *et al.*, 1994; BIGGAR *et al.*, 1996). A segunda, Pós-Colombiana, especula que o HTLV-1 tenha sido disseminado da África para o Novo Mundo e até para o Japão, durante o tráfico de escravos, no período compreendido entre os séculos XVI a XIX (MIURA *et al.*, 1994; VAN DOOREN *et al.*, 1998; YAMASHITA *et al.*, 1999).

Durante a colonização portuguesa, entre os séculos XVI e XIX, em torno de 4 milhões de pessoas foram trazidas da África para o Brasil, como resultado do tráfico de escravos, principalmente para as regiões nordeste e sudeste. Apesar da maioria dos africanos trazidos para a Bahia, durante o tráfico de escravos, ter vindo do oeste da África, especificamente do Benin e da Nigéria, Rodrigues (1977) mostra evidências de que africanos também foram trazidos de outras regiões do sul do continente africano, atualmente conhecidas como Angola, África do Sul e Moçambique (VERGER, 1968; RODRIGUES, 1977; VIANA FILHO, 1988).

Estudos filogenéticos baseados em análises do gene *env* e da região LTR permitiram classificar o HTLV-1 em sete subtipos: HTLV-1a ou Cosmopolita (MIURA *et al.*, 1994; 1997); HTLV-1b ou Centro Africano (HAHN *et al.*, 1984; VANDAMME *et al.*, 1994); HTLV-1c ou oriundo da Melanésia

(GESSAIN *et al.*, 1992); HTLV-1d, isolado de pigmeus em Camarões e no Gabão (CHEN *et al.*, 1995; MAHIEUX *et al.*, 1997); HTLV-1e, isolado de pigmeus na República Democrática do Congo (SALEMI *et al.*, 1998); HTLV-1f, de um indivíduo do Gabão e HTLV-1g, recentemente descrito como um novo subtipo em Camarões, na África Central (WOLFE *et al.*, 2005). O subtipo Cosmopolita reúne cepas de diferentes regiões geográficas, sendo o mais disseminado no mundo, de maneira que é subdividido em 5 subgrupos, a depender de sua localização geográfica: A – Transcontinental, B – Japonês, C – Oeste Africano, D – Norte Africano e E – Negro do Peru.

Estudos de caracterização molecular do HTLV-1 demonstram a predominância do subtipo Cosmopolita em infecções no Brasil, sendo a maioria dos isolados pertencentes ao subgrupo Transcontinental (SEGURADO *et al.*, 2002; ALCANTARA *et al.*, 2003), e em menor proporção isolados que caracterizam o subgrupo Japonês (YAMASHITA *et al.*, 1999).

## 1.6 Doenças associadas ao HTLV-1

A TSP/HAM caracteriza-se por uma mielopatia lentamente progressiva, tendo como consequência a paraparesia dos membros inferiores (ROMAN & OSAME, 1988). Anormalidades sensoriais periféricas, dor lombar persistente, hiperreflexia, disfunção da bexiga e do intestino e impotência sexual são outros sinais relacionados à doença. A prevalência da TSP/HAM é de 1 a 5% em pacientes infectados pelo vírus (KAPLAN *et al.*, 1990).

Na TSP/HAM, a patogênese tem relação com a desmielinização local atribuída à invasão da célula infectada no Sistema Nervoso Central, e o desencadeamento de uma resposta inflamatória crônica. Três principais hipóteses tentam explicar a patogênese da TSP/HAM (IJICHI *et al.*, 1993; TAYLOR, 1998; NAGAI *et al.*, 2000; JACOBSON, 2002; OSAME, 2002). A primeira pressupõe que

o HTLV-1 infecta células da glia as quais apresentariam antígenos virais em sua superfície celular. Os linfócitos T citotóxicos (CTL) específicos circulantes atravessariam a barreira hemato-encefálica, e encontrariam a célula infectada causando sua morte por liberação de citocinas. A segunda hipótese presume que um antígeno próprio na célula da glia seja similar ao antígeno viral. Desta forma, células T CD4<sup>+</sup> encontrariam este antígeno viral na periferia e ao atravessar a barreira hemato-encefálica confundiriam a célula da glia com uma célula infectada disparando uma resposta auto-imune com morte da célula da glia. Na terceira hipótese, células T CD4<sup>+</sup> infectadas com o HTLV-1 e linfócitos T CD8<sup>+</sup> específicos anti-HTLV migrariam através da barreira hemato-encefálica, se encontrariam no sistema nervoso central e as células da glia seriam destruídas pelas citocinas liberadas pelos CTLs contra as células T CD4<sup>+</sup> infectadas.

A ATL, por sua vez, é um linfoma/leucemia agressivo, associada à infecção pelo HTLV-1, que tem seu curso clínico classificado em quatro estágios: aguda, crônica, linfoma e *smouldering* (SHIMOYAMA, 1991). Além desses tipos clínicos, existe ainda outra categoria conhecida como ATL cutânea, cujas manifestações são restritas à pele, e que subsequenteemente foi subdividida em dois outros subtipos: tumoral e eritematosa (BITTENCOURT *et al.*, 2007). ATL caracteriza-se pela infiltração de células T CD4<sup>+</sup> malignas nos linfonodos, baços, trato gastro-intestinal e pele, além da presença de células T anormais com núcleo lobulado ou forma de flor (*flower cells*) (MATSUOKA, 2005). A proporção de ocorrência de ATL em homens e mulheres infectados no Japão é de 6% e 2%, respectivamente, sendo o período médio entre a infecção inicial e o desfecho clínico de 60 e 40 anos para infectados no Japão e na Jamaica, respectivamente (YASUNAGA & MATSUOKA, 2007).

Na patogênese pelo HTLV-1, a proteína regulatória viral, Tax, funciona como um agente fundamental no desenvolvimento das diferentes patologias. Além de regular a expressão de genes virais, tax interage com fatores de transcrição celulares (CREB/ATF, NF- $\kappa$ B e SRF) e moléculas de

sinalização para estimular ou reprimir a expressão de genes celulares. 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. Além dessa atividade transativadora, Tax 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).

### **1.7 Resposta Imune na Infecção pelo HTLV-1**

A resposta imune em combate às infecções virais pode ser segregada em dois grandes pólos: um de natureza antígeno-inespecífica e outro de caráter antígeno-específica. No que se refere ao segundo pólo, ao entrar em contato com proteínas virais, os linfócitos B interagem com estruturas antigênicas dos vírus através dos receptores de superfície dos linfócitos. Essas moléculas são capazes de reagir internalizando os determinantes antigênicos e dando início ao processo de ativação dos linfócitos T CD4+. Após a interação com os antígenos virais e conseqüente maturação em plasmócitos, anticorpos são produzidos para eliminação do antígeno por neutralização ou opsonização.

Neste contexto, existem evidências do envolvimento da resposta humoral nos eventos que acompanham a evolução da infecção crônica pelo HTLV-1 (LAL *et al.*, 1993). A maioria dos anticorpos produzidos em resposta à infecção natural pelo HTLV-1 inclui aqueles direcionados contra as glicoproteínas do envelope viral: gp46 e gp21. Desta forma, os anticorpos específicos contra determinantes antigênicos do envelope viral podem inibir a formação do sincício sugerindo uma possível proteção contra o ciclo de infecção (HADLOCK *et al.*, 2002).



O HTLV-1 infecta preferencialmente células T periféricas, predominantemente linfócitos TCD4<sup>+</sup> de memória (CD45RO) e, em menor frequência, linfócitos T CD8<sup>+</sup>, observando-se inicialmente um padrão policlonal de integração viral.

A resposta imune do hospedeiro frente à infecção viral, principalmente a resposta celular desencadeada por células T CD8<sup>+</sup> específicas anti-HTLV, é reconhecida como um evento crucial, determinando o rumo da infecção. Estudos recentes têm sugerido que esta resposta celular é influenciada pela via de infecção do hospedeiro, mucosa ou sangue periférico, além de fatores genéticos individuais como polimorfismos em genes de HLA (*Human Leucocitary Antigen*) e genes envolvidos na resposta imune. A resposta antiviral através de linfócitos T citotóxicos é ativada *in vivo* e direcionada principalmente contra epítomos específicos na proteína regulatória Tax do HTLV, e em menor proporção aos epítomos em Gag, Env e Pol (KANNAGI *et al.*, 1991; PARKER *et al.*, 1992; ELOVAARA *et al.*, 1993; PIQUE *et al.*, 1996). Desta forma, dependendo do epítomo viral e dos fatores genéticos do hospedeiro, tem se estabelecido a existência de reservatórios distintos de CTLs em resposta aos diferentes epítomos, com a mesma função de reconhecimento de células infectadas, via HLA/peptídeo, no entanto, com proliferação e liberação de citocinas diferenciadas (LIM *et al.*, 2000). Bangham & Osame (2005) demonstraram que a taxa de lise, via CTL, de células T CD4<sup>+</sup> expressando tax foi correlacionada negativamente com a carga proviral tanto em indivíduos assintomáticos como em indivíduos TSP/HAM.

A ativação das células T CD4<sup>+</sup>, no entanto, parece ser um evento precoce na infecção pelo HTLV-1, já que indivíduos infectados, porém assintomáticos, possuem elevado percentual de células T CD4<sup>+</sup> HLA-DR<sup>+</sup>, não sendo esse um bom indicador de progressão clínica para TSP/HAM. Segundo Goon e col. (2004), a resposta CD4<sup>+</sup> tanto em indivíduos assintomáticos como em indivíduos

TSP/HAM é direcionada preferencialmente contra as proteínas do envelope, em contraste à resposta imunodominante de linfócitos T CD8<sup>+</sup> específicos contra a proteína Tax.

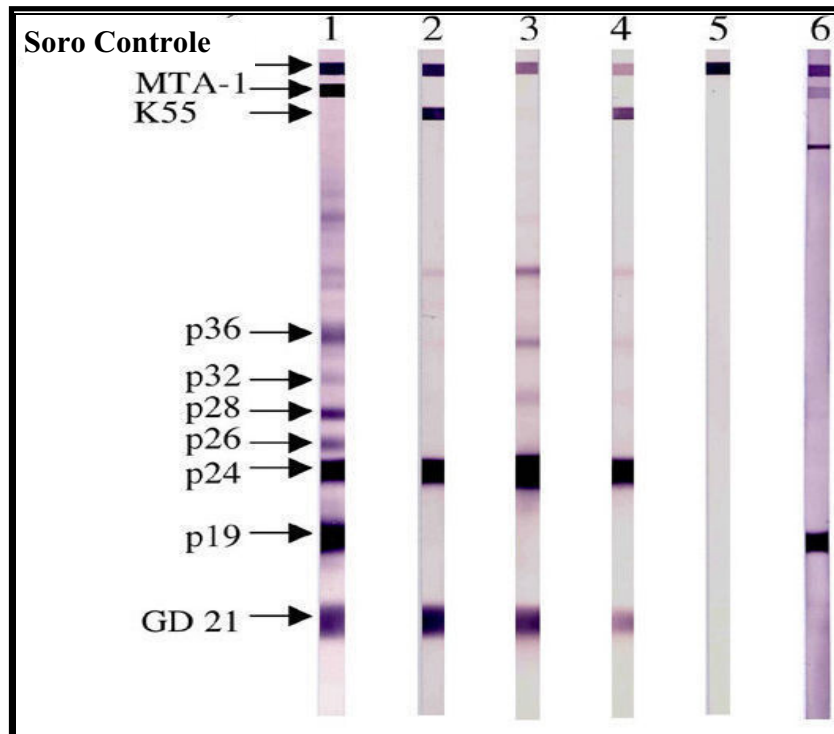
## 1.8 Diagnóstico sorológico

Partindo do pressuposto de que a maioria das proteínas virais são imunogênicas e que, portanto, anticorpos são produzidos para reagir contra elas, estes são detectados no soro de pessoas infectadas com o HTLV, o que torna viável a existência dos testes sorológicos para diagnóstico, como *ELISA* e *Western Blot* (WB) (CONSTATINE *et al.*, 1992).

O diagnóstico sorológico da infecção pelo HTLV baseia-se na detecção de anticorpos específicos contra o vírus. Os métodos sorológicos podem ser classificados em duas categorias: os testes de triagem e os de confirmação. Os ensaios de triagem detectam anticorpos contra o HTLV-1 e HTLV-2, porém ao serem usados em população de baixo risco como doadores de sangue, o valor preditivo positivo pode ser muito baixo, sendo necessária a confirmação do resultado por meio dos ensaios confirmatórios, que detêm maior especificidade e que podem também discriminar a presença de anticorpos específicos contra o HTLV-1 e HTLV-2. Proteínas estruturais codificadas pelos genes *gag* e *env* têm importância no reconhecimento laboratorial da infecção. Os primeiros testes foram introduzidos utilizando lisado viral como única fonte antigênica, e mais recentemente, ensaios baseados em antígenos recombinantes ou peptídeos sintéticos foram introduzidos isoladamente ou em combinação com o lisado viral. O teste mais utilizado na triagem sorológica do HTLV é o ensaio imunoenzimático (*Enzyme Immuno Assay*-EIA), onde 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 humana marcado. Para a etapa de confirmação, o ensaio mais utilizado é o teste WB (LAL *et al.*,

1992), que permite 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). Como ainda não existe um consenso internacional para a interpretação dos testes de WB, a Organização Mundial de Saúde (OMS) adotou um critério que inclui a reatividade para uma proteína do gene *gag* (p19 e p24) e para uma proteína do envelope viral (gp46 e gp21), para que uma amostra seja considerada positiva (**Fig. 7**). Algumas amostras podem ter resultado positivo no WB, sem diferenciação entre os tipos 1 e 2 sendo consideradas portanto, positivas mas não tipadas. Quando uma amostra reage com uma das bandas, porém não completa o critério de positividade, seu resultado é considerado indeterminado (**Fig. 7**). Apesar de ser um teste confirmatório, estudos têm demonstrado que enquanto o WB pode detectar facilmente anticorpos contra produtos do gene *gag*, ele, freqüentemente, falha na detecção de anticorpos contra os produtos do gene *env*. Por isso, na maioria das vezes, para que se obtenha um diagnóstico mais seguro, é necessária uma combinação de testes, agregado ao WB, o que contribui para o alto custo e para demora na conclusão do diagnóstico (LAL *et al.*, 1992).

Algumas explicações, como imunossupressão em indivíduos co-infectados HTLV-1/HIV ou HTLV-1/2, são sugeridas para entender o fenômeno da soroideterminação. Outras razões para explicar o perfil indeterminado no WB seriam a soroconversão, especialmente em populações de risco ou em áreas endêmicas, ou a infecção por cepas virais divergentes e/ou defectivas (THORSTENSSON *et al.*, 2002).



**Fig. 7- Representação esquemática dos diferentes resultados no *Western Blot*.** Fita 1: amostra HTLV-1 positiva. Fitas 2 e 4: amostras HTLV-2 positiva. Fita 3: amostra não tipada. Fita 5: amostra negativa. Fita 6: amostra indeterminada.

## 2. JUSTIFICATIVA

A infecção pelo HTLV-1 é endêmica em diferentes regiões geográficas do mundo (HANCHARD *et al.*, 1990; MUELLER, 1991; GESSAIN & DE THE, 1996), sendo Salvador a cidade do Brasil com a mais alta prevalência da infecção na população geral (1,8%), e a região amazônica conhecida por ser área endêmica pela infecção pelo HTLV-2 (DOURADO *et al.*, 2003; ISHAK *et al.*, 2003).

Grande parte dos trabalhos sobre a epidemiologia do HTLV-1 consiste em estudos de soroprevalência em doadores de sangue, no entanto, há uma relativa escassez de estudos de caracterização molecular desses isolados virais. É importante caracterizar genotipicamente os isolados de HTLV-1 de diferentes regiões geográficas do Brasil, inclusive com contribuições étnicas distintas, para obter informações mais detalhadas sobre as possíveis rotas migratórias do vírus para o país e dentro do território nacional. Afinal, existem pelo menos duas hipóteses que articulam possíveis origens do vírus na América do Sul, seja por uma introdução Pré-Colombiana ou Pós-Colombiana.

Além dos estudos de epidemiologia molecular, que se baseiam em análises filogenéticas utilizando, majoritariamente, a região LTR, demais estudos de investigação dos produtos gênicos do vírus podem fornecer direcionamentos para o desenvolvimento de vacinas terapêuticas.

Neste aspecto, o gene *env* tem importância elementar, já que a maioria dos anticorpos produzidos em resposta à infecção natural pelo HTLV-1 inclui aqueles direcionados contra as glicoproteínas do envelope: gp46 e gp21. Sabe-se inclusive que a região central da gp46 é responsável por 90% dos anticorpos específicos que são produzidos nos indivíduos infectados (SHERMAN *et al.*, 1993). Desta forma, parece lógico a condução de estudos que tentam canalizar a

produção de anticorpos específicos para o controle da expansão das células infectadas, bem como para o controle de fenômenos como a formação de sincício (HADLOCK *et al.*, 2002).

Outro aspecto importante no curso natural da infecção é o reconhecimento de partículas virais ou mesmo células infectadas por linfócitos especificamente ativados que irão promover o estabelecimento dos mecanismos efetores. A ativação dos linfócitos T, por sua vez, é dependente da apresentação de peptídeos virais associados à moléculas de HLA, na superfície da célula apresentadora de antígeno.

Respostas CTL e T CD4+ contra o HTLV-1 são componentes importantes da resposta imune no decorrer e no controle da infecção, além do que o escape viral de reconhecimento por CTL pode ter relação com a patogênese do vírus, assim como acontece em infecções pelo HIV-1, sugerindo que a seleção de mutantes deve ser dirigida, pelo menos em parte, pela pressão seletiva mediada pela resposta CTL.

Acreditamos, então, que o desenvolvimento de uma possível vacina terapêutica deve ser baseado na utilização de alvos virais eficientes, bem como em estratégias que tenham o objetivo de manipular a resposta imune para o controle da disseminação de partículas virais.

Entendemos, por fim, que ferramentas de bioinformática representam um grande aliado na análise do grande número de seqüências virais, bem como na geração de dados moleculares relevantes para o direcionamento da realização de etapas mais amplas e onerosas no caminho longo, mas necessário, do desenvolvimento de estratégias terapêuticas eficazes.

### 3. OBJETIVOS

#### 3.1 Objetivo Geral

Caracterização molecular dos isolados do HTLV-1 para dar suporte às hipóteses de migrações das populações humanas infectadas, bem como aos estudos de escolha de candidatos vacinais.

#### 3.2 Objetivos Específicos

- Realizar levantamento e montagem de banco de dados com todos os isolados do HTLV-1, referentes aos genes *env* e *pol*, previamente publicados no banco mundial de seqüências (GenBank);
- Identificar, nas seqüências previamente publicadas dos genes *env* e *pol*, epítomos ligantes de alelos de HLA classes I e II;
- Caracterizar as seqüências previamente publicadas do gene *env*, de acordo com parâmetros físico-químicos como: antigenicidade, flexibilidade, acessibilidade e hidrofobicidade;
- Identificar, nas seqüências previamente publicadas do gene *env*, possíveis sítios de modificação pós-traducional;
- Identificar, nas seqüências previamente publicadas do gene *env*, a presença de sítios (mutações) selecionados positivamente;
- Padronizar a utilização das ferramentas de bioinformática destinadas à predição de epítomos, cálculo de pressão seletiva, análise físico-química e identificação de domínios potenciais;
- Contribuir, com estudos moleculares dos genes *env* e *pol*, no desenvolvimento de estratégias para o desenho de protótipos vacinais, além de fornecer dados, sobre o gene *env*, que sejam úteis a estratégias de melhoria do diagnóstico sorológico confirmatório;

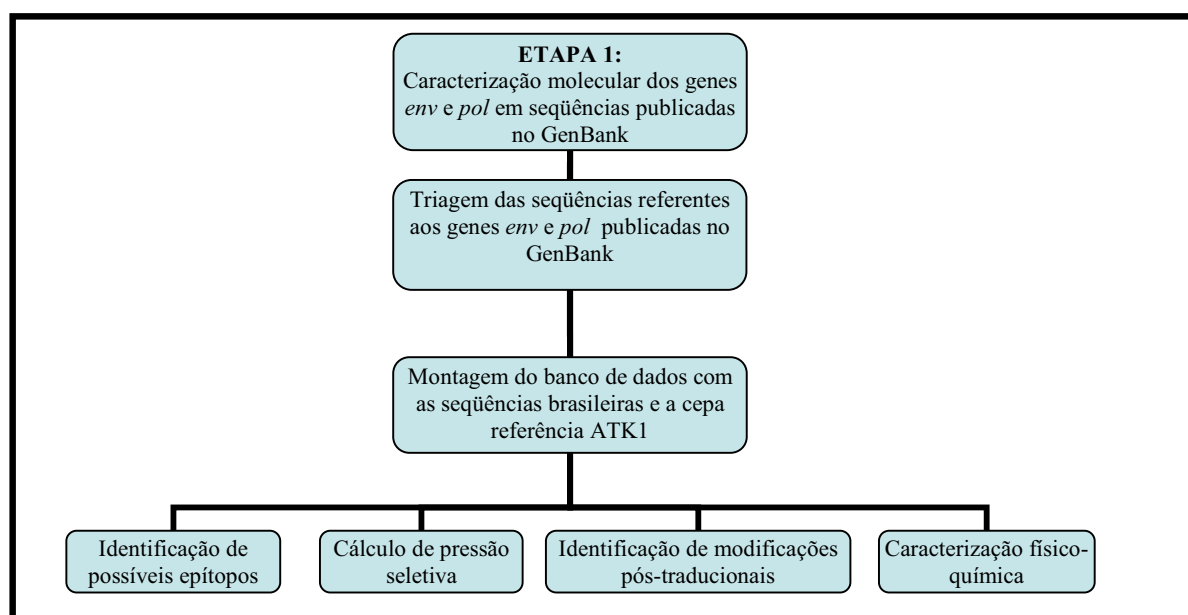
- Estimar a soroprevalência do HTLV-1 em doadores de sangue atendidos no HEMOACRE (Rio Branco-Acre), em Julho de 2004;
- Caracterizar, genotipicamente, isolados do HTLV-1 identificados neste estudo e em estudos anteriores de soroprevalência, conduzidos nesta população (Rio Branco-Acre);
- Discutir aspectos epidemiológicos relacionados à infecção pelo vírus nesta população, e em infecções em Salvador, Fortaleza e no Peru.



## 4. METODOLOGIA

### 4.1 Caracterização molecular de seqüências dos genes *env* e *pol*, disponíveis no Banco Mundial de seqüências

**4.1.1-Delineamento do estudo:** na primeira etapa foi realizado um levantamento, no GenBank (Banco Mundial de Seqüências), de todas as seqüências referentes aos genes *env* e *pol* publicadas a partir de isolados de HTLV-1 no Brasil, para a caracterização molecular. As seqüências foram então organizadas em um banco de dados, que foi submetido à identificação de epítomos, caracterização físico-química, identificação de modificações pós-traducionais e identificação de sítios selecionados positivamente (**Fig. 8**).



**Fig. 8-** Delineamento do estudo de caracterização molecular dos genes *env* e *pol* em seqüências publicadas no GenBank.

**4.1.2-Characterização dos objetos de estudo:** para a realização do mapeamento molecular das seqüências protéicas referentes aos genes *env* e *pol* originadas de isolados virais brasileiros, um banco de dados foi organizado com todas as seqüências alvo que tivessem sido, previamente, publicadas e submetidas ao banco mundial de seqüências GenBank. A organização do banco de dados revelou, então, existir 15 seqüências referentes ao gene *env*, e 43 seqüências referentes ao gene *pol*. As 15 seqüências *env* correspondem ao fragmento que compreende o nucleotídeo 535 da gp46 ao nucleotídeo 153 da gp21, enquanto que as seqüências *pol* correspondem à região que começa no nucleotídeo 2.058 até o nucleotídeo 2.375. Todas as localizações nucleotídicas referem-se à cepa referência ATK1 (SEIKI *et al.*, 1983).

- Critérios de inclusão: 1-Ser uma seqüência originada de um isolado viral, identificado em infecção no Brasil, 2-Ser referente ao gene *env*, ou ao gene *pol*, e 3-Estar disponível no GenBank.
- Critérios de exclusão: 1- Ter tamanho menor que 500pb para o gene *env*, ou menor que 200pb para o gene *pol*.

**4.1.3-Identificação de possíveis epítomos lineares:** a procura para a existência de possíveis epítomos ligantes de 14 alelos de HLA-I (HLA A 01, HLA A 08, HLA A 0201, HLA A 0702, HLA A 2402, HLA A 2709, HLA A 5101, HLA A 6801, HLA A 4402, HLA A 2705, HLA A 1510, HLA A 0203, HLA A 26, HLA A 03) e 6 alelos de HLA-II (HLA DRB1 0301, HLA DRB1 0701, HLA DRB1 1501, HLA DRB1 0101, HLA DRB1 0401, HLA DRB1 1101), foi realizada utilizando uma ferramenta *online* denominada SYFPEITHI (RAMMENSEE *et al.*, 1999) (<http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm>). Esta ferramenta fornece

informações sobre a seqüência do epítopo, o alelo de HLA, ao qual o epítopo é específico, os aminoácidos de ancoramento na fenda da molécula de HLA e o *score* de ligação de cada epítopo ao alelo do HLA estudado.

**4.1.4- Caracterização físico-química:** as seqüências peptídicas do gene *env* foram submetidas à análise físico-química através do NPSA (*Network Protein Sequence Analysis*) (HOPP & WOODS, 1983) (<http://npsa-pbil.ibcp.fr/>), para conhecimento sobre o perfil de antigenicidade, hidrofobicidade, acessibilidade e flexibilidade, como ferramenta de apreciação da resposta imune mediada por anticorpos e no reconhecimento de peptídeos apresentados na superfície da célula, via associação com moléculas de HLA.

**4.1.5- Identificação de domínios potenciais:** nesta etapa, ao submeter as seqüências peptídicas do gene *env*, especialmente a região dos epítotos, à ferramenta Prosite (FALQUET *et al.*, 2002) implementada no programa GeneDoc (NICHOLAS *et al.*, 1997) foi realizada uma busca por assinaturas específicas conhecidas como domínios potenciais protéicos, que caracterizam-se por sítios susceptíveis às modificações pós-traducionais, que podem ser adição de resíduos para fosforilação de quinases, glicosilação ou ainda clivagem proteolítica da ligação peptídica.

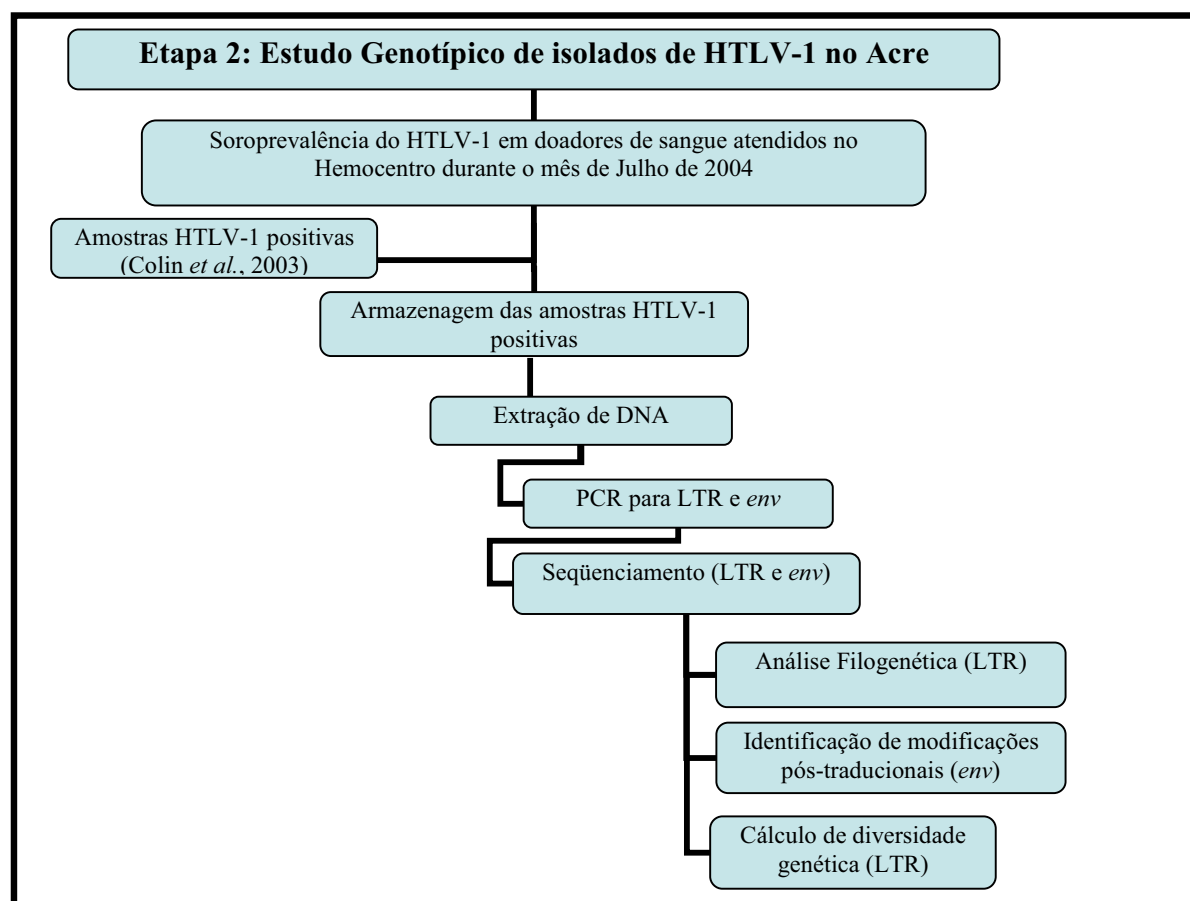
**4.1.6- Identificação das mutações:** as variações nucleotídicas, assim como as trocas nas seqüências de aminoácido foram identificadas, após o alinhamento das seqüências no Clustal X (JEANMOUGIN *et al.*, 1998), através da utilização dos programas de edição de seqüências BioEdit (HALL, 1999) e GeneDoc (NICHOLAS *et al.*, 1997).

**4.1.7-Cálculo de pressão seletiva:** para testar a hipótese de que substituições nucleotídicas, no gene *env*, poderiam ser resultado de eventos de seleção natural, a pressão seletiva positiva foi testada, calculada para cada códon, utilizando seis diferentes modelos de substituição (M0, M1, M2, M3, M7 e M8) baseados na máxima verossimilhança (YANG *et al.*, 2000). Para tanto, adotamos os pressupostos evolutivos do Neodarwinismo, que admitem que mutações randômicas resultam em variação genética, sobre a qual age a seleção natural, alterando a frequência dos alelos na população, representando, portanto, uma força evolutiva dominante. Todos os modelos testados estão disponíveis no CODEML, implementado no PAML (YANG, 1997). De acordo com estudos prévios (DE-OLIVEIRA *et al.*, 2004), os valores de  $\omega$  e  $p$  são estimados de acordo com a otimização dos valores de verossimilhança, de forma que sítios com valores de probabilidade excedendo 90% e valor de  $\omega$  superior a 1 são classificados como sítios selecionados positivamente. Por fim, o cálculo da razão de verossimilhança (*Likelihood Ratio Test-LRT*) foi usado para determinar se: I) existe heterogeneidade entre os sítios e II) se há sítios selecionados positivamente.

## 4.2 Estudo genotípico e epidemiológico de isolados do HTLV-1 de Rio Branco-Acre

**4.2.1-Delineamento do estudo:** na segunda etapa foi realizado um estudo de soroprevalência do HTLV-1 em doadores de sangue atendidos no Hemocentro de Rio Branco-Acre, triados durante o mês de Julho de 2004. Tanto as amostras HTLV-1 positivas provenientes deste estudo (AC181), como as amostras HTLV-1 positivas provenientes de um estudo anterior (AC042, AC069, AC129, AC174, AC204 e AC057) de soroprevalência (COLIN *et al.*, 2003) foram submetidas à PCR para a região LTR e para o gene *env*. Após seqüenciamento e, portanto geração das seqüências consenso, os fragmentos da região LTR foram, portanto, submetidos à análise filogenética e ao cálculo de diversidade genética para

caracterização genotípica dos isolados, enquanto que as seqüências do gene *env* foram incluídas na caracterização molecular através da identificação de modificações pós-traducionais (Fig. 9).



**Fig. 9-** Delineamento do estudo genotípico e epidemiológico de isolados do HTLV-1 de Rio Branco-Acre.

**4.2.2- Caracterização da população:** o estudo de soroprevalência para a infecção pelo HTLV-1 foi conduzido durante o mês de Julho de 2004, tendo feito parte da casuística, 219 doadores de sangue, atendidos no Hemocentro de Rio Branco-Acre, que foram triados sorologicamente, para a presença de anticorpos específicos. O estudo de caracterização genotípica foi conduzido com a única amostra HTLV-1 positiva (AC181) deste estudo de prevalência, e com seis amostras HTLV-1 positivas

(AC042, AC069, AC174, AC57, AC204 e AC129), previamente armazenadas no Laboratório Avançado de Saúde Pública (LASP), provenientes de um estudo prévio de soroprevalência (COLIN *et al.*, 2003), nesta mesma população.

Dados do IBGE (Instituto Brasileiro de Geografia e Estatística), 2006, relatam que a população do estado de Acre se caracteriza, etnicamente, por cerca de 66,5% de pardos, 26% de brancos, 6,8% de pretos, e 0,7% de indígenas. Dados obtidos por meio de pesquisa de autodeclaração.

**4.2.3-Sorologia para HTLV-1/2:** todas as 219 amostras dos doadores de sangue do Hemocentro do Acre foram submetidas ao diagnóstico sorológico. Este consistiu de uma triagem inicial caracterizada por um ensaio imunoenzimático (EIA) (ORTHO<sup>®</sup> HTLV-1/2 *Ab-Capture* ELISA *test system*, Ortho-Clinical Diagnostics, Raritan, EUA). As amostras com resultados inconclusivos ou indeterminados por EIA, foram analisadas por *Western Blot* (HTLV BLOT 2.4, Genelabs Diagnostics<sup>®</sup>, Singapore). Todos os procedimentos de reação, bem como os resultados obtidos foram conduzidos e avaliados de acordo com as especificações do fabricante.

**4.2.4-Extração de DNA:** o DNA genômico foi extraído a partir de amostras de sangue total, utilizando o kit de extração Qiagen (*QIAamp<sup>®</sup> DNA Blood Kit*), sendo a concentração final de DNA medida por espectrofotometria (*Gene Quantpro RNA/DNA Calculator*).

**4.2.5-Reação em cadeia da polimerase (PCR):** foi realizada PCR para o gene *env* e para a região LTR. O gene *env* foi amplificado utilizando os métodos previamente padronizados (YANG *et al.*,1997). Enquanto, a região LTR foi amplificada utilizando dois pares de *primers* responsáveis pela amplificação de dois fragmentos em sobreposição: um segmento correspondente à

região 5'LTR-*gag* de 473 bp e um outro segmento da região *tax*-3'LTR de 479 bp, como previamente descrito (ALCANTARA *et al.*, 2006).

**4.2.6-Seqüenciamento:** os produtos de PCR amplificados foram purificados usando o kit de purificação QIAGEN (QIAquick® PCR Purification Kit) e seqüenciados no seqüenciador automático ABI3100 utilizando o kit *Taq FS Dye Terminator Cycle Sequencing Kit* (Applied Biosystems). As reações de seqüenciamento foram conduzidas em duplicata e com utilização dos mesmos *primers* utilizados na PCR. A seqüência consenso, gerada a partir dos fragmentos, *forward* e *reverse*, gerados no seqüenciamento, bem como a avaliação da qualidade do seqüenciamento foram realizadas utilizando o programa SeqScape (Applied Biosystems SeqScape Software v 2.5).

**4.2.7-Análise filogenética:** a análise filogenética, baseada na região LTR, incluiu seqüências referência de diferentes regiões geográficas e grupos étnicos distintos, selecionadas do *NCBI/Nucleotide Sequence Database-GenBank*, e que representassem todos os subtipos e subgrupos descritos. Ao final das análises, todas as seqüências, geradas neste trabalho, foram submetidas ao GenBank. As seqüências foram alinhadas usando o *software* Clustal X (JEANMOUGIN *et al.*, 1998) e editadas manualmente no programa GeneDoc (NICHOLAS *et al.*, 1997). O modelo evolutivo Tamura Nei (TrN + G), que leva em consideração diferentes taxas de substituição nucleotídica para transições e transversões, assim como taxa de substituição heterogênea entre os sítios, foi selecionado através do *software* Modeltest (POSADA & CRANDALL, 1998), como sendo o melhor modelo para o *data set* utilizado. As reconstruções baseadas nos métodos *Neighbor-Joining* (NJ) e *Maximum-Likelihood* (ML) foram geradas utilizando o *software* PAUP\* 4.0b10 (SWOFFORD, 1998). A árvore NJ foi construída com otimização dos parâmetros: matriz de substituição nucleotídica e distribuição gama com parâmetro

$\alpha=0.811083$ . A reprodutibilidade da topologia NJ foi assegurada através da análise de *Bootstrap* com 1000 réplicas. O teste da razão de verossimilhança foi usado para calcular o suporte estatístico do tamanho dos ramos: sendo o valor de  $p < 0,001$ , considerado altamente significativo (\*\*), e valor de  $p < 0,05$ , considerado significativo (\*). Suportes de *Bootstrap* e ML foram incluídos à árvore NJ editada no TreeView 1.4 (PAGE, 1996). A genotipagem das seqüências LTR parciais foi realizada utilizando a ferramenta online: *LASP HTLV-1 Automated Genotyping Tool* (<http://lasp.cpqgm.fiocruz.br>).

**4.2.8-Cálculo da diversidade genética:** com o objetivo de conhecer mais sobre a origem do HTLV-1 no país, foram calculadas as diversidades genéticas, entre as seqüências LTR total geradas (AC181 e AC042), neste trabalho, e em comparação com outras seqüências LTR originadas de isolados de diferentes regiões geográficas do Brasil (Salvador, Fortaleza) e do Peru, utilizadas na reconstrução filogenética: IDUSSA, HB3203, HB3135, FNN 19, FNN153, FNN158, HB3229, HB3114, FNN100, HB3311, FNN156 de Salvador; MASU, FCR, JCP e MAQS de Fortaleza; PE15bo, Bl2, Qu1, PE8si, PE7hu, Qu2, Qu3, PE6fa, Me2, PE13sa, Me1, PE14ru do Peru. Os cálculos de diversidade genética foram feitos utilizando o modelo de distância Tamura Nei, com suporte estatístico, desvio padrão e 1000 réplicas através da análise de *Bootstrap* implementado no pacote de programas do MEGA 3.0 (KUMAR *et al.*, 1994).

**4.2.9-Identificação de domínios potenciais:** as seqüências do gene *env* foram submetidas à ferramenta Prosite (FALQUET *et al.*, 2002) implementada no programa GeneDoc (NICHOLAS *et al.*, 1997) para busca por assinaturas específicas conhecidas como domínios potenciais protéicos, que caracterizam-se por sítios susceptíveis às modificações pós-traducionais, que podem ser



adição de resíduos para fosforilação de quinases ou glicosilação ou ainda clivagem proteolítica da ligação peptídica.

**4.2.10-Aspectos éticos:** este projeto foi aprovado pelo Comitê de ética em Pesquisa do Instituto de Saúde Coletiva-Universidade Federal da Bahia- ISC/UFBA. Todos os doadores de sangue foram informados sobre os procedimentos e condutas quanto à coleta do material biológico, bem como a utilização dele, e concordaram em assinar o termo de consentimento livre e esclarecido (TCLE) para utilização de suas amostras neste projeto.

## 5. RESULTADOS

### 5.1 Caracterização molecular de seqüências dos genes *env* e *pol*, previamente publicadas

As análises de predição de epítomos permitiram identificar um epítomo em cada peptídeo estudado, demonstrando que os peptídeos relacionados são alvos importantes para o reconhecimento do sistema imune, através dos ligantes específicos dos alelos de HLA-I/II. Com o objetivo de investigar a especificidades dos epítomos, calculamos a similaridade na composição de aminoácidos entre as seqüências brasileiras, desta forma, observamos que o número de mutações encontradas para o tamanho do fragmento, resultou numa variação de 9% e 17% para epítomos específicos de HLA-I e HLA-II, respectivamente. A diferença entre a variação protéica total de epítomos HLA-I e HLA-II específicos, pode ser explicada pelo fato de que o tamanho típico de um ligante de HLA-I é de 9 aminoácidos, enquanto que o comprimento de um ligante HLA-II é muito maior (15 aa). Todos os epítomos mapeados foram capazes de se ligar a maioria dos alelos de HLA estudados, o que é importante como estratégia de vencer a dificuldade do polialelismo em genes de HLA, para o desenvolvimento de uma vacina.

A análise físico-química demonstrou que todo o fragmento gp46-gp21 é caracterizado por alta antigenicidade, flexibilidade e hidroflicidade. Além disso, foi possível identificar que mutações presentes em 5 dos 12 epítomos foram responsáveis pelo aumento da antigenicidade na região.

Todos os resultados estão organizados no artigo publicado no *Memórias do Instituto Oswaldo Cruz*.

**5.1.1- Artigo publicado:** *Mem Inst Oswaldo Cruz* 102(6): 741-749, 2007

**Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design**

Aline Cristina Mota-Miranda<sup>\*</sup>, Tulio de-Oliveira<sup>\*\*</sup>, Domingos Ramon Moreau, Catarina Bomfim<sup>\*\*\*</sup>, Bernardo Galvão-Castro<sup>\*</sup>, Luiz Carlos Junior Alcantara<sup>\*/+</sup>

Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Moniz-Fiocruz, Salvador, BA, Brasil <sup>\*</sup>Escola Bahiana de Medicina e Saúde Pública, Fundação Bahiana para o Desenvolvimento das Ciências, Salvador, BA, Brasil <sup>\*\*</sup>MRC Pathogen Bioinformatics Unit, South African National Bioinformatics Institute, UWC, Cape Town, South Africa and Zoology Department, Oxford University, Oxford, United Kingdom <sup>\*\*\*</sup>Faculdade de Tecnologia e Ciência, Salvador, BA, Brasil

Key words: Human T-cell Lymphotropic Virus type 1; Env; Pol; epitope

Financial support: Lasp/CPqGM/Fiocruz, Fapesb (grant 303/03), Brazilian Ministry of Health (grant 306/04 and 307/04)

+Corresponding author: lalcan@cpqgm.fiocruz.br

Luiz Carlos Junior Alcantara, Ph D.

Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz. Rua Waldemar Falcão 121, Candeal, Salvador, Bahia, Brasil.

40296-610.

Tel: # 55 71 31762255 - Fax # 55 71 3176 2300

This study was carried out to evaluate the molecular pattern of all available Brazilian human T-cell lymphotropic virus type 1 Env (n = 15) and Pol (n = 43) nucleotide sequences via epitope prediction, physico-chemical analysis, and protein potential sites identification, giving support to the Brazilian AIDS vaccine program. In 12 previously described peptides of the Env sequences we found 12 epitopes, while in 4 peptides of the Pol sequences we found 4 epitopes. The total variation on the amino acid composition was 9 and 17% for human leukocyte antigen (HLA) class I and class II Env epitopes, respectively. After analyzing the Pol sequences, results revealed a total amino acid variation of 0.75% for HLA-I and HLA-II epitopes. In five of the twelve Env epitopes the physico-chemical analysis demonstrated that the mutations magnified the antigenicity profile. The potential protein domain analysis of Env sequences showed the loss of a CK-2 phosphorylation site caused by D197N mutation in one epitope, and a N-glycosylation site caused by S246Y and V247I mutations in another epitope. Besides, the analysis of selection pressure have found eight positive selected sites ( $w = 9.59$ ) using the codon-based substitution models and maximum-likelihood methods. These studies underscore the importance of this Env region for the virus fitness, for the host immune response and, therefore, for the development of vaccine candidates.

**Key words:** human T-cell lymphotropic virus type 1; Env; Pol; epitope

The human t-cell lymphotropic virus type 1 (HTLV-1) is the etiologic agent of adult T-cell leukemia lymphoma (ATLL) (Poiesz et al. 1980) and tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM) (Gessain et al. 1985, Osame et al. 1986). This infection is endemic in Japan, in the Caribbean Basin, some South American, and African regions (Mueller 1991), while there are 2.5 million HTLV-1 infected people in Brazil (Carneiro-Proietti et al. 2002). In contrast, only 5% of infected people develop the associated disease, and it is not yet known why 95-98% of them remain asymptomatic. The envelope glycoproteins of this virus are known to play a critical role in the infection process. The surface glycoprotein (gp46) subunit is involved in cellular receptor recognition, while the transmembrane glycoprotein (gp21) subunit anchors the gp46 to the cell and plays a major role in the post-binding steps of the fusion process, resulting in the formation of multinucleated giant cells named syncytia. These envelope proteins are expressed on the surface of virus-infected cells and on viral particles and they are the first to be recognized by the host in the course of the natural immune response (Nagy et al. 1983, Palker et al. 1989), in such a way that HTLV-1 infected individuals develop a strong immune response to the envelope (*env*) gene products. Several studies have focused on mapping linear immunodominant regions of the HTLV-1 surface glycoprotein that could elicit antibody responses, for the purpose of vaccine development and diagnostic screening (Tanaka et al. 1991, Inoue et al. 1992, Kuroki et al. 1992, Baba et al. 1993, Desgranges et al. 1994). Specific antibodies against envelope determinants can inhibit HTLV-1-mediated syncytium formation, suggesting possible protection from the infection cycle (Hadlock et al. 2002). Multiple neutralizing monoclonal antibodies to linear epitopes within amino acids 175-200 of the gp46 and more rarely to epitopes into the carboxy terminal of the protein have been described using immunoreactive assays (Baba et al. 1993). The antiviral cytotoxic T-lymphocyte (CTL) is activated in vivo and directed mainly to the

HTLV-1 Tax regulatory protein and to a lesser extent to the structural Gag, Env, and Pol proteins (Kannagi et al. 1991, Parker et al. 1992, Elovaara et al. 1993, Pique et al. 1996).

Possible strategies for vaccine development should evaluate the antigenic capacity of viral peptides in stimulating the humoral and cellular immune responses. Therefore, a vaccine development process would be more specific if the practice of screening important genomic regions would be carried out. This study was therefore developed to evaluate the molecular pattern of all available Brazilian HTLV-1 Env and Pol sequences concerning to: physico-chemical analysis, to infer the peptide capacity of generating functional antibodies; protein potential site analysis, to get information of a possible immune system escape mechanism, such as changes in the protein structure; and epitopes prediction, to identify the possible targets for B and T-cells through human leukocyte antigen (HLA) molecules presentation. Furthermore, it is important to make sure that bioinformatic tools are safe and precise for screening new peptides to be tested *in vitro*.

## MATERIALS AND METHODS

*Sequences selection* - All sequences analyzed in this study were selected from the GenBank, between March and May 2006. There were 44 available *pol* gene nucleotide sequences in the GenBank, but one of them (U12108) were rejected from the analysis because it was a too short (140pb) fragment. At that time also, there were about 59 available *env* gene nucleotide sequences, but 44 of them were rejected from the analysis because they were isolated from Argentine, Caribbean, Romanian, and South African infected individuals, and because they were fragments with less than 500 bp. So, we have analyzed fifteen *env* gene sequences, corresponding to the nucleotide 535 of gp46 to 153 of the gp21, and 43 *pol* sequences, from nucleotide 2,058 to 2,375, that represent all *env* and *pol* virus strains isolated from Brazilian individuals until May, 2006. To perform all analyses, two alignments (*env* and

*pol*) were carried out with the ATK1 reference strain (Seiki et al. 1983) using the Clustal X 1.83 software (Jeanmougin et al. 1998).

All *env* nucleotide sequences previously deposited in the GenBank and used in the study are listed below with their corresponding accession number: PT9ATL (U81869); PT12ATL (U81865); PT3ATL (U81866); PT5ATL (U81867); PT8ATL (U81868). The new nucleotide sequences are: *env* - IDUSSA (DQ007198); FNN159 (DQ007197); FNN155 (DQ007194); FNN148 (DQ007191); FNN149 (DQ007197); FNN153 (DQ007193); FNN158 (DQ007158); FNN09 (DQ007189); FNN100 (DQ007190); FNN156 (DQ007195); *pol* - G0110 (AF197327); TP98/70 (AF197326); TP98/66 (AF197325); TP98/55 (AF197324); TP97/38 (AF197323); TP97/35 (AF197322); TP97/33 (AF197321); RNT97/38 (AF197320); FNN094 (AF197319); FNN091 (AF197318); FNN089 (AF197317); FNN087 (AF197316); FNN083 (AF197315); FNN082 (AF197314); FNN081 (AF197313); FNN080 (AF197312); FNN078 (AF197311); FNN073 (AF197310); FNN072 (AF197309); FNN071 (AF197308); FNN068 (AF197307); FNN064 (AF197306); FNN061 (AF197305); FNN060 (AF197304); FNN057 (AF197303); FNN054 (AF197302); FNN053 (AF197301); FNN051 (AF197300); FNN047 (AF197299); FNN043 (AF197298); FNN041 (AF197297); FNN039 (AF197296); FNN035 (AF197295); FNN032 (AF197294); FNN029 (AF197293); FNN028 (AF197292); FNN026 (AF197291); FNN022 (AF197290); FNN019 (AF197289); FNN009 (AF197288); G0263 (AF197287); G0153 (AF197286); TP98/75 (AF197285).

*Epitopes prediction* - Before performing the epitope prediction, we have screened from the literature, twelve and four peptides previously described and tested *in vitro* (Horal et al. 1991, Baba et al. 1993, Pique et al. 1996, Schonbach et al. 1996), corresponding to the *env* and *pol* gene, respectively. The main criterion used to screen the peptides were their frequency in the papers. The epitope prediction was carried out for all Env and Pol amino acid sequences to 14 HLA-I (HLA A26, HLA

A1510, HLA A4402, HLA A01, HLA A0201, HLA A2402, HLA A5101, HLA A03, HLA A0203, HLA A2705, HLA A6801, HLA A08, HLA A0702, HLA A2709 ) and 6 HLA-II alleles (HLA DRB1 0101, HLA DRB1 0401, HLA DRB1 0301, HLA DRB1 1501, HLA DRB1 0701, HLA DRB1 1101), using the online bioinformatics tool SYFPEITHI (Rammensee et al. 1999) ([http://www.syfpeithi.de/Scripts/MHCServer.dll/Epitope Prediction.htm](http://www.syfpeithi.de/Scripts/MHCServer.dll/Epitope%20Prediction.htm)). This tool uses an algorithm that can predict sequences that have the potential ability to bind to one or more different HLA-I and HLA-II molecules. It also provides information about the epitope sequence, the specificity to the HLA molecule and the HLA binding score for each epitope. We performed the epitope prediction to the HLA alleles available in the software which are the most frequent in the database. After epitope prediction, similarity and variation calculations were performed comparing the peptide and epitope amino acid sequences within the Brazilian sequences, to get information about the specificity of the epitope binding to the HLA molecule and a possible epitope consensus sequence among different isolates.

*Physico-chemical and potential protein domain analysis* - To investigate possible influences of mutations, described into the *env* gene epitopes, we performed the physico-chemical analysis of the Env sequences using the Network Protein Sequence Analysis (NPSA) (Argos et al. 1982, Kyte & Doolittle 1982, Hopp & Woods 1983, Karplus & Schulz 1985, Parker et al. 1986) (<http://npsa-pbil.ibcp.fr/>) and the potential protein domain analysis using the GeneDoc software (Nicholas et al. 1997) and the Prosite tool, as previously described (Queiroz et al. 2007).

*Selective pressure test* - To test the hypothesis that the amino acid substitutions within the predicted epitopes into the *env* gene could have been favored or not by natural selection, the positive selection was assessed using six different codon-based maximum-likelihood (ML) substitution models (Yang et al. 2000). All models were implemented in the Codeml program of the PAML software



package (Yang 1997). In agreement with a previously described study (de Oliveira et al. 2004), the  $\omega$  and  $p$  values were estimated through maximum-likelihood optimization, in such a way that using the M3 model, sites with a posterior probability exceeding 90% and a  $\omega$  value  $> 1$  were designated as being “positive selection sites”. Finally, likelihood ratio test (LRT) analysis was used to determine: (1) if site heterogeneity selection was present and (2) if there were to be positive selection sites.

## RESULTS

*Epitopes prediction* - Based on the peptide screening study, we identified the presence of 12 previously described peptides in all Brazilian Env sequences. These peptides cover the entire studied gp46 fragment and their respective locations in the protein according to ATK1 reference sequence are listed below (Table I). The total variation between the published peptide amino acid sequence and its composition in the Brazilian isolates was 17%, confirming that this viral protein region has too conservative characteristics. To search for possible epitopes into the peptides, we used SYFPEITHI tool. From our results, it was possible to identify one epitope in each studied peptide, demonstrating that the screened peptides are true targets for host immune response, through the recognition of an HLA-I or HLA-II specific binder. To investigate the epitopes specificity, among viral isolates, we have calculated the similarity of the amino acid epitope sequence in all Brazilian studied strains. Our results indicated that the observed mutations resulted in a total variation of 9 and 17% for HLA-I and HLA-II specific epitopes, respectively. The difference between the total variation of HLA-I and HLA-II epitopes can be explained by the fact that the typical length of HLA-I ligand comprises 9aa, while the length of an HLA-II ligand is much longer (15aa). Interestingly, both Env and Pol predicted epitopes, through the computer analysis, were able to bind more than one HLA allele, demonstrating a

promiscuous behavior of those antigenic determinants and also suggesting an approach of these results to the biological significance.

A similar set of analyses such as peptide screening, epitopes prediction, and similarity/variation calculations were also carried out for the 43 Pol amino acid sequences. Only the four most frequent peptides previously described (Pique et al. 1996, Schonbach et al. 1996) were screened to the Pol fragment analysis because the analyzed Brazilian sequences are just 172 amino acids longer. In such a way, all four published peptides were found in all Brazilian isolates and revealed conservative characteristics, through the amino acid peptide variation of 0.75%, which was much lower than the *env* gene percentages (Table II). When we performed the epitope prediction on the Pol peptides sequences, we found four HLA-I specific epitopes and four HLA-II specific epitopes, both of them with a very low rate of total epitope variation (0.75%) (Table II).

*Physico-chemical and potential protein domain analysis* - The physico-chemical analysis demonstrated that the entire studied fragment (gp46-gp21) is characterized by high antigenicity, flexibility, and hydrophilicity profiles. Interestingly, the results demonstrated that the mutations present in five of twelve epitopes had increased the antigenicity profile. When subjecting these epitopes to potential protein domain analysis to investigate if the mutations in the epitope had changed any signature into the protein (Table I), we observed that the D197N mutation in the first epitope (195-207aa) was the cause for the loss of a CK2- phosphorylation site (Fig. 1A). Similarly, the tenth epitope (238-250aa) which was characterized by both S246Y and V247I mutations did not show the N-glycosylation site present in the epitope without these mutations (Fig. 1B).

*Selective pressure test* - ML methods were used to assess amino acid variation and identify targets of positive selection. The fraction of sites ( $p_1, p_2, p_3$ ), with the respective  $\omega$  ( $dn/ds$ ) values for

each model (M0-M3, M7-M8) are presented in the Table III. Although the env gene is considered to be too conservative, it was possible to identify eight sites (codons) which are positive selected ( $\omega = 9.59$ ) using models M2-M3 and M8. The LRT comparing the models support these results ( $p < 0,005$ ).

## DISCUSSION

In the present study, we have identified a set of multiple HLA-A and HLA-DR binding CD8 and CD4 T-cell epitopes, respectively, derived from the computer analysis performed at the SYFPHEITHI tool. The identified epitopes are conserved motifs into the envelope protein of viral isolates, since the variation of amino acid in the epitopes sequences was less than 20%. The knowledge about the epitope sequence variation has great importance because the similarity between the predicted epitopes could be very useful to design an efficient vaccine for most common viral isolates. Furthermore, other studies using sera from HTLV-1-infected individuals have identified mutations within amino acids 175-200 that can affect antibody-mediated neutralization of HTLV-1 gp46-mediated syncytium formation (Blanchard et al. 1999, Tallet et al. 2001). Moreover, the observation of which alleles could bind to the mutant epitope is important, because the epitope containing the wild anchor motif could bind better to HLA molecules than other mutant motifs. Identifying which HLA allele the epitope is specific for could be also very useful to direct vaccine design for a specific population whose most frequent alleles have the strongest epitope's binder. Moreover, the computer epitope prediction is already used to the engineering of an HIV vaccine, confirming the utility of bioinformatics tools to select and construct novel "immunogenic consensus sequence" T cell epitopes for a globally relevant vaccine against HIV (De Groot et al. 2005, Fonseca et al. 2006).

The epitope prediction performed in this study is based on comparisons of precursor peptide sequences known to contain epitopes. It is the discovery of allele-specific motifs shared by eluted natural MHC ligands that allow the exact prediction of peptides from a given protein sequence. Every HLA allele has its individual peptide specificity that is defined by the position and specificity of the anchoring and non-anchoring amino acid residues. SYFPEITHI uses motifs matrices deduced from refined motifs based on a given pool of sequences and a single peptide analysis to search for natural ligands.

Although, these data had been obtained in computer analysis, our results are in agreement with experimental studies that have selected Tax and Env peptides to induce secretion of gamma interferon in peripheral blood mononuclear cells obtained from monkeys chronically infected with HTLV-1. After immunization, a high titre of antibodies and a high frequency of IFN- $\delta$ -producing cells were detected against the Env and the Tri-Tax immunogens, but not to the individual Tax peptides (Kazanji et al. 2006).

In line with previous reports, the post-transcriptional modifications, especially the glycosylation events are related to the formation of neutralizing antibodies, what explain our intention to compare the epitopes prediction results with carbohydrate moieties. Besides, there is an effort to engineer conformationally dependent HTLV-1 epitopes that elicit a protective immune response, especially a specific immune recognition and antibody responses (Kaumaya et al. 1995).

The physico-chemical analysis demonstrated that the entire studied fragment (gp46-gp21) had high antigenicity, flexibility and hidrophilicity profiles, and that mutations in that region could be important to the conformational epitope. This strategy is already investigated by previous studies, which report that peptide vaccines, to be able to induce high affinity and protective neutralizing

antibodies, must rely in part on the design of antigenic epitopes, that mimic the three-dimensional structure of the corresponding region in the native protein (Sundaram et al. 2004).

It is known that a single-nucleotide mutation into an immunodominant CTL epitope can lead to viral escape and increased viral replication in HIV studies. Although there are no studies referring this experience with HTLV-1 isolates, two sites under positive selective pressure in our analysis were also responsible for the loss of CK2-phosphorylation and N-glycosylation moieties. This finding is very important to underscore the importance of the *env* gene to the immune response, and to confirm that this region is hardly submitted to the selective pressure, and also to the viral fitness.

Concerning that the proteins encoded by *pol* gene are not strong immune response inducers and that they are neither under positive selection nor an antibody response target, the Pol peptides were not submitted to physico-chemical analysis nor potential domains.

This study was conducted with all available viral sequences isolated from Brazilian infected individuals with the objective to obtain a national panel of this kind of molecular investigation. In fact, there are not many Brazilian sequences of these regions available in the GenBank, but at least all of information here showed about epitope prediction, physico-chemical and protein potential site analysis, are new and could give support to future studies. We have not analyzed tax protein, because there were only five available sequences in the GenBank at May 2006, and they corresponded to a short genomic region, but next studies should evaluate the tax protein, because the HTLV-1 tax protein is therefore a source of naturally processed epitopes presented by both HLA-I and HLA-II molecules, and are capable of stimulating both CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses directed against tumors *in vivo* (Hanabuchi et al. 2001).

Nowadays, bioinformatics tools are being used to screen some immunodominant regions before testing *in vitro*, to get information about the genetic dynamics of the virus isolates and to monitor

epidemic changes, such as their implications into the treatment response. Therefore, computer-aided analysis of protein sequences using various correlates of protein antigenicity is an important tool to the study of antigenic determinants.

### **Acknowledgment**

To Elisabeth Deliege for the technical assistance.

### **REFERENCES**

- Argos P, Rao JK, Hargrave PA 1982. Structural prediction of membrane-bound proteins. *Eur J Biochem* 128: 565-575
- Baba E, Nakamura M, Tanaka Y, Kuroki M, Itoyama Y, Nakano S, Niho Y 1993. Multiple neutralizing B-cell epitopes of human T-cell leukemia virus type 1 (HTLV-1) identified by human monoclonal antibodies. A basis for the design of an HTLV-1 peptide vaccine. *J Immunol* 151: 1013-1024
- Blanchard S, Astier-Gin T, Tallet B, Moynet D, Londos-Gagliardi D, Guillemain B 1999. Amino acid changes at positions 173 and 187 in the human T-cell leukemia virus type 1 surface glycoprotein induce specific neutralizing antibodies. *J Virol* 73: 9369-9376
- Carneiro-Proietti AB, Ribas JG, Catalan-Soares BC, Martins ML, Brito-Melo GE, Martins Filho OA, Pinheiro SR, Araújo AQ, Galvao-Castro B, de Oliveira MS, Guedes AC, Proietti FA 2002. Infection and disease caused by the human T cell lymphotropic viruses type I and II in Brazil. *Rev Soc Bras Med Trop* 35: 499-508

- De Groot AS, Marcon L, Bishop EA, Rivera D, Kutzler M, Weiner DB, Martin W 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine* 23: 2136-2148
- De Oliveira T, Salemi M, Gordon M, Vandamme AM, Van Rensburg EJ, Engelbrecht S, Coovadia HM, Cassol S 2004. Mapping Sites of Positive Selection and Amino Acid Diversification in the HIV Genome: An Alternative Approach to Vaccine Design? *Genetics* 167: 1047-1058
- Desgranges C, Souche S, Vernant JC, Smadja D, Vahlne A, Horal P 1994. Identification of novel neutralization-inducing regions of the human T cell lymphotropic virus type I envelope glycoproteins with human HTLV-I-seropositive sera. *AIDS Res Hum Retroviruses* 10: 163-173
- Elovaara I, Koenig S, Brewah AY, Woods RM, Lehky T, Jacobson S 1993. High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. *J Exp Med* 177: 1567-1573
- Fonseca SG, Coutinho-Silva A, Fonseca LAM, Segurado AC, Moraes SL, Rodrigues H, Hammer J, Kallás EG, Sidney J, Sette A, Kalil J, Cunha-Neto E 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by the HIV-1 infected patients. *AIDS* 20: 2263-2273
- Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de The G 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* 2: 407-410
- Hadlock KG, Yang Q, Rowe J, Fong SK 2002. Epitope mapping of human monoclonal antibodies recognizing conformational epitopes within HTLV type 1 gp46, employing HTLV Type 1/2 envelope chimeras. *AIDS Res Hum Retroviruses* 18: 57-70

- Hanabuchi S, Ohashi T, Koya Y, Kato H, Hasegawa A, Takemura F, Masuda T, Kannagi M 2001. Regression of human T-cell leukemia virus type I (HTLV-I)-associated lymphomas in a rat model: peptide-induced T-cell immunity. *J Natl Cancer Inst* 93: 1775-1783
- Hopp TP, Woods KR 1983. A computer program for predicting protein antigenic determinants. *Mol Immunol* 20: 483-489
- Horal P, Hall WW, Svennerholm B, Lycke J, Jeansson S, Rymo L, Kaplan MH, Vahlne A 1991. Identification of type-specific linear epitopes in the glycoproteins gp46 and gp21 of human T-cell leukemia viruses type I and type II using synthetic peptides. *Proc Natl Acad Sci* 88: 5754-5758
- Inoue Y, Kuroda N, Shiraki H, Sato H, Maeda Y 1992. Neutralizing activity of human antibodies against the structural protein of human T-cell lymphotropic virus type I. *Int J Cancer* 52: 877-880
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ 1998. Multiple sequence alignment with clustal X. *Trends Biochem Sci* 23: 403-405
- Kannagi M, Harada S, Maruyama I, Inoko H, Igarashi H, Kuwashima G, Sato S, Morita M, Kidokoro M, Sugimoto M *et al.* 1991. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8<sup>+</sup> cytotoxic T cells directed against HTLV-I-infected cells. *Int Immunol* 8: 761-767
- Karplus PA, Schulz GE 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72: 212-213
- Kaumaya PT, Conrad SF, DiGeorge AM, Lairmore MD 1995 Glycosylation-dependent peptide antigenic determinants of env gp46 HTLV-1. *Leukemia* 9(Supp 1): 133-8



- Kazanji M, Heraud JM, Merien F, Pique C, Thé G, Gessain A, Jacobson S 2006 Chimeric peptide vaccine composed of B- and T-cell epitopes of human T-cell leukemia virus type 1 induces humoral and cellular immune responses and reduces the proviral load in immunized squirrel monkeys (*Saimiri sciureus*). *J Gen Virol* 87: 1331-1337
- Korber B 1997. HIV Signature and Sequence Variation Analysis. In AG Rodrigo, GH Learn *Computational Analysis of HIV Molecular Sequences*, Kluwer Academic Publisher, Dordrecht, Netherlands, p. 55-72
- Kuroki M, Nakamura M, Itoyama Y, Tanaka Y, Shiraki H, Baba E, Esaki T, Tatsumoto T, Nagafuchi S, Nakano S *et al.* 1992. Identification of new epitopes recognized by human monoclonal antibodies with neutralizing and antibody dependent cellular cytotoxicity activities specific for human T cell leukemia virus type 1. *J Immunol* 149: 940-948
- Kyte J, Doolittle RF 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105-132
- Mueller N 1991. The epidemiology of HTLV-I infection. *Cancer Causes Control* 2: 37-52
- Nagy K, Clapham P, Cheingsong-Popov R, Weiss RA 1983. Human T-cell leukemia virus type I: induction of syncytia and inhibition by patients' sera. *Int J Cancer* 32: 321-328
- Nicholas KB, Nicholas HBJ, Deerfield DW 1997. GeneDoc: Analysis and visualization of genetic variation. *EMB News* 14: 30
- Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M 1986. HTLV-I associated myelopathy, a new clinical entity. *Lancet* 1: 1031-1032

Palker TJ, Tanner ME, Scarce RM, Streilein RD, Clark ME, Haynes BF 1989. Mapping of immunogenic regions of human T cell leukemia virus type I (HTLV-I) gp46 and gp21 envelope glycoproteins with env-encoded synthetic peptides and a monoclonal antibody to gp46. *J Immunol* 142: 971-978

Parker JM, Guo D, Hodges RS 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25: 5425-5432

Parker CE, Daenke S, Nightingale S, Bangham CR 1992. Activated, HTLV-I-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. *Virology* 188: 628-636

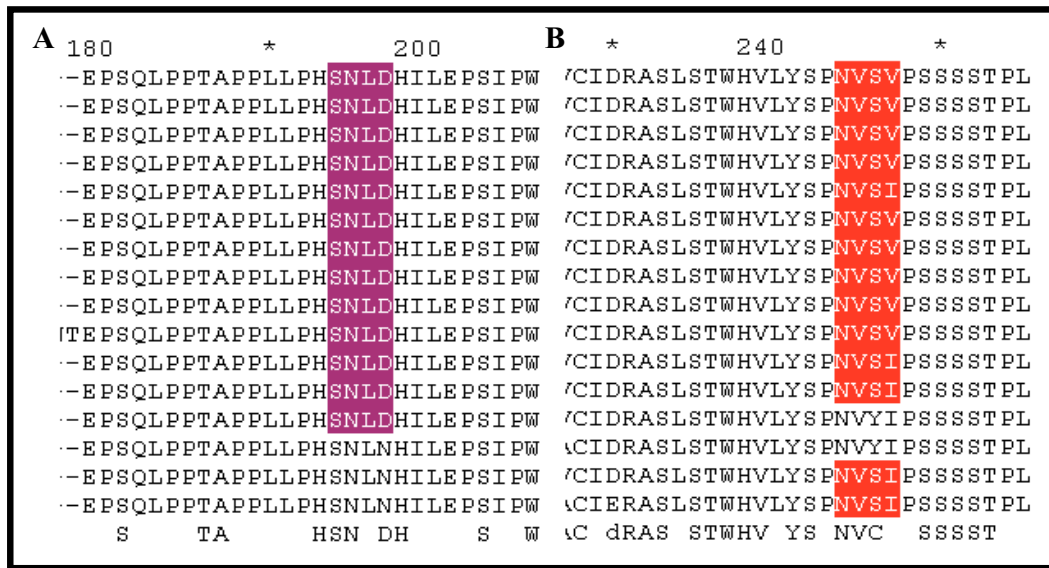
Pique C, Connan F, Levilain JP, Choppin J, Dokh elar MC 1996. Among all human T- cell leukemia virus type 1 proteins, tax, polymerase, and envelope proteins are predicted as preferential targets for the HLA-A2 restricted cytotoxic T-cell response. *J Virol* 70: 4919-4926

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* 77: 7415-7419

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

- Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50: 213-219
- Schönbach C, Nokihara K, Bangham CR, Kariyone A, Karaki S, Shida H, Takatsu K, Egawa K, Wiesmüller KH, Takiguchi M 1996. Identification of HTLV-1-specific CTL directed against synthetic and naturally processed peptides in HLA-B\* 3501 transgenic mice. *Virology* 226: 102-112
- Seiki M, Hattori S, Hirayama Y, Yoshida M 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci* 80: 3618-3622
- Sundaram R, Lynch MP, Rawale SV, Sun Y, Kazanji M, Kaumaya PT 2004. De novo design of peptide immunogens that mimic the coiled coil region of human T-cell leukemia virus type-1 glycoprotein 21 transmembrane subunit for induction of native protein reactive neutralizing antibodies. *J Biol Chem* 279: 24141-51
- Tallet B, Astier-Gin T, Moynet D, Londos-Gagliardi D, Guillemain B 2001. Sequence variations in the amino- and carboxy-terminal parts of the surface envelope glycoprotein of HTLV type 1 induce specific neutralizing antibodies. *AIDS Res Hum Retroviruses* 17: 337-348
- Tanaka Y, Zeng L, Shiraki H, Shida H, Tozawa H 1991. Identification of a neutralization epitope on the envelope gp46 antigen of human T cell leukemia virus type I and induction of neutralizing antibody by peptide immunization. *J Immunol* 147: 354-360
- Yang Z 1997. PAML: a program package for phylogenetic analysis by maximum likelihood . *Comput Appl Biosci* 13: 555-556

Yang Z, Nielsen R, Goldman N, Pedersen AM 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449



**Fig. 1:** potential protein domain analysis: loss of CK2- phosphorylation (A) and N-glycosylation (B) sites because of the D197N and (S246Y and V247I) mutations, respectively.



**Table II**

Human T-cell lymphotropic virus type I previous published Pol peptides versus possible human leukocyte antigen class I and II epitopes and similarity with Brazilian sequences.

MHC	Peptide sequence	Similarity %	SYFPEITHI epitope prediction		
			Epitope sequence	Similarity (%)	HLA allele
I	704-S L L Q A I A Y L-712 ..... 43/43 <sup>a</sup>	100	702-I S S L L Q A I A Y L G K-714 ..... 43/43 <sup>b</sup>	100	▲
	777-L P M D N A L S I-785 . S ..... 1/43 <sup>a</sup> . I ..... 1/43 <sup>a</sup>	97	775-P D L P M D N A L S I A L-787 ..... S ..... 1/43 <sup>b</sup> ..... I ..... 1/43 <sup>b</sup>	97	2/3/4/6/7/8/10 11/12/13
	709-I A H L G K P S Y-717 ..... 43/43 <sup>a</sup>	100	707-Q A I A H L G K P S Y I N-719 ..... 43/43 <sup>b</sup>	100	▲
	738-L A I R H T T H V P Y-748 ..... 43/43 <sup>a</sup>	100	736-T S L A I R H T T H V P Y N P-750 ..... 43/43 <sup>b</sup>	100	▲
	Total peptide variation into the sequences:0,75%		Total epitope variation into the sequences: 0,75%		
II	704-S L L Q A I A Y L-712 ..... 43/43 <sup>a</sup>	100	702-I S S L L Q A I A Y L G K P S Y I N T-720 ..... 43/43 <sup>b</sup>	100	▲▲
	777-L P M D N A L S I-785 . S ..... 1/43 <sup>a</sup> . I ..... 1/43 <sup>a</sup>	97	775-P D L P M D N A L S I A L W T I N-791 ..... S ..... 1/43 <sup>b</sup> ..... I ..... 1/43 <sup>b</sup>	97	2/3/4/5/6/7/8 10/11/12/13
	709-I A H L G K P S Y-717 ..... 43/43 <sup>a</sup>	100	707-Q A I A H L G K P S Y I N T D N G P A-725 ..... 43/43 <sup>b</sup>	100	▲▲
	738-L A I R H T T H V P Y-748 ..... 43/43 <sup>a</sup>	100	736-T S L A I R H T T H V P Y N P T S S G L V-750 ..... 43/43 <sup>b</sup>	100	▲▲
	Total peptide variation into the sequences:0,75%		Total epitope variation into the sequences: 0,75%		

<sup>a</sup> : Peptide mutation frequency in all Brazilian sequences (n=43); <sup>b</sup> : Epitope mutation frequency in all Brazilian sequences (n=43); HLA I alleles: 1(HLA A 1510), 2 (HLA A 5101), 3 (HLA A 0201), 4 (HLA A 03), 5 (HLA A 01), 6 (HLA A 6801), 7 (HLA A 026), 8 (HLA 2705), 9 (HLA A 2402), 10 (HLA A 2709), 11 (HLA A 4402), 12 (HLA A 0702), 13 (HLA A 08) and 14 (HLA A 0203); HLA II alleles: 1 (HLA DRB1 0101), 2 (HLA DRB1 1101), 3 (HLA DRB1 0701), 4 (HLA DRB1 0301), 5 (HLA DRB1 1501) and 6 (HLA DRB1 0401); ▲ represent all HLA-I and ▲▲ HLA-II alleles.

**Table III**

Human T-cell lymphotropic virus type 1 Env selective pressure test: parameter estimated under six models of variable  $\omega$  ( $d_n/d_s$ ) among sites.

	<b>M0</b>	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M7</b>	<b>M8</b>	<b>LRT I</b>	<b>LRT II</b>
<b><i>env</i> gene</b>	lnL= -864.96 w = 0.72	lnL= -863.14 p <sub>1</sub> = 0.61 p <sub>2</sub> = 0.39 w <sub>1</sub> = 0.0 w <sub>2</sub> = 1.0	lnL= -859.36 p <sub>1</sub> = 0.91 p <sub>2</sub> = 0.0 p <sub>3</sub> = 0.08 w <sub>1</sub> = 0.0 w <sub>2</sub> = 1.0 w <sub>3</sub> = 9.59	lnL = -859.36 p <sub>1</sub> = 0.91 p <sub>2</sub> = 0.0 p <sub>3</sub> = 0.08 w <sub>1</sub> = 0.0 w <sub>2</sub> = 1.0 w <sub>3</sub> = 9.59	lnL= -863.81 p <sub>1</sub> = 0.33 p <sub>2</sub> = 0.33 p <sub>3</sub> = 0.33 w <sub>1</sub> = 0.0 w <sub>2</sub> = 1.0 w <sub>3</sub> = 1.0	lnL= -859.36 p <sub>1</sub> = 0.30 p <sub>2</sub> = 0.30 p <sub>3</sub> = 0.30 p <sub>4</sub> = 0.08 w <sub>1</sub> = 0.0 w <sub>2</sub> = 0.0 w <sub>3</sub> = 0.0 w <sub>4</sub> = 9.59	11.2 (p< 0.05)	7.56 (p< 0.05)
<b>Selected sites (P = 1.0)</b>			197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P	197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P		197 N 214 G 216 I 227 A 230 E 246 S 247 I 288 P		

The location of the selected codons is described according to HTLV-1 ATK1 reference sequence. LTR results support the presence of positive selection sites (p<0.05) in the *env* gene.



## 5.2 Caracterização genotípica de novos isolados do HTLV-1 de Rio Branco:

A prevalência de HTLV-1 entre doadores de sangue, atendidos no Hemocentro de Rio Branco-Acre, em Julho de 2004 foi de 0,46%, já que foi possível detectar apenas um caso de infecção em 219 indivíduos triados.

Desse único caso de infecção, apenas foi possível amplificar a região LTR total. Em contrapartida, das seis amostras HTLV-1 positivas originadas do estudo de caracterização sorológica anterior (COLIN *et al.*, 2003), que estavam armazenadas no LASP, apenas foi possível gerar 7 seqüências, sendo uma da região LTR total (AC042), 3 fragmentos LTR parciais (AC069, AC174 e AC129) e três seqüências do gene *env* (ENVAC57, ENVAC69 e ENVAC204).

A análise filogenética dos dois fragmentos completos da região LTR (AC181 e AC042) permitiu classificá-los como membros do subgrupo Transcontinental do subtipo Cosmopolita, com valor de *Bootstrap* de 71% e  $p < 0,001$  para análise de ML. Outras três seqüências (AC069, AC129 e AC174) correspondentes à região LTR parcial, por serem fragmentos menores (350pb), foram submetidas à ferramenta de genotipagem, sendo também classificadas como pertencentes ao subgrupo Transcontinental do subtipo Cosmopolita.

Todos os resultados, referentes a este estudo, estão organizados no manuscrito aceito para publicação no *Journal of Medical Virology*.

### 5.2.1- Artigo Publicado: *J Medical Virology*, 80(11):1966-1971, 2008

#### **HTLV-1 infection in blood donors from Western Brazilian Amazon region: seroprevalence and molecular study of viral isolates**

Authors: Aline Cristina Mota-Miranda<sup>±</sup>, Sérgio Pereira Araújo<sup>±</sup>, Juarez Pereira Dias<sup>~</sup>, Denise Duizit Colin<sup>#</sup>, Simone Kashima<sup>§</sup>, Dimas Tadeu Covas<sup>§</sup>, José Tavares-Neto<sup>~</sup>, Bernardo Galvão-Castro<sup>±,¶</sup>, Luiz Carlos Junior Alcantara<sup>±,¶</sup>.

*Author's affiliation:* <sup>±</sup> Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Muniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brasil; <sup>#</sup>Fundação Hospital Estadual o Acre, Rio Branco, Acre, Brasil, <sup>§</sup>Hemocentro de Ribeirão Preto, São Paulo, Brasil; <sup>~</sup>Universidade Federal da Bahia, Salvador, Bahia, Brasil <sup>¶</sup>Escola Bahiana de Medicina e Saúde Pública, Fundação Bahiana para o Desenvolvimento das Ciências, Salvador, Bahia, Brasil.

Institution where the analysis were performed:

*Laboratório Avançado de Saúde Pública, Centro de Pesquisa Gonçalo Muniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brasil*

Corresponding author:

Luiz Carlos Junior Alcantara

Advanced Public Health Laboratory, Gonçalo Moniz Research Center, Oswaldo Cruz Foundation.

Waldemar Falcão Street 121, Candeal, Salvador, Bahia, Brazil – 40296-610.

Telephone # 55 71 31762255

Fax # 55 71 31762300

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

*Shortened title:* **HTLV-1 in the Western Brazilian Amazon region**

To contribute further with the HTLV-1 seroprevalence in Brazil, and to discuss the virus molecular epidemiology in this Amazon population (Rio Branco-Acre), 219 blood donors were screened for HTLV-1-specific antibodies. It was detected only a single case of infection (0.46% seroprevalence) among blood donors screened during July 2004 at the FUNDACRE. To assess molecular epidemiology, Neighbor-joining and Maximum Likelihood phylogenetic analyses of two (n=2) complete LTR region sequences were performed with PAUP\* software. Since the HTLV-1 envelope surface (gp46) and transmembrane (gp21) glycoproteins are important to virus fitness, three available envelope glycoproteins sequences (n=3) were analyzed using the Prosite tool to determinate potential protein sites. Phylogenetic analysis demonstrated that the single new isolate, described in this study, and also the unpublished LTR strain described in a previous report belong to the Transcontinental subgroup of the Cosmopolitan subtype, inside the Latin American cluster A. Similar result was obtained when submitting, to the Automated Genotyping System, three LTR partial sequences from a previous study of HTLV-1 seroprevalence in the same Amazon population. The potential protein site analysis showed, that all *env* sequences encoded two PKC phosphorylation sites at amino acid (aa) positions 310-312 and 342-344, one CK2 phosphorylation site at 194-197aa, three N-glycosylation sites at 222-225aa, 244-247aa and 272-275aa, and a single N-myristylation site at 327-338aa. In conclusion, potential protein sites described in HTLV-1 gp46 and gp21 confirm the presence of conserved sites in HTLV-1 envelope proteins, likewise phylogenetic analysis suggest a possible recent introduction of the virus into North Brazil.

**Key-words:** LTR region, glycoproteins, phylogenetic analysis, protein sites.

## INTRODUCTION

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

According to HTLV-1 LTR region analysis, six genetic subtypes have been proposed [Miura et al., 1994]: a or Cosmopolitan (worldwide distribution); b or Central African; c or Melanesian; d, from Central African pygmies, Cameroon and Gabon; e, from Democratic Republic of Congo; f, isolated initially from an infected individual from Democratic Republic of Congo, and also described in Cameroon; and g, recently described as a new subtype in Cameroon. The Cosmopolitan subtype is

divided in five different subgroups: A-Transcontinental, B- Japanese, C-West African/Caribbean, D-North African and E-Black Peruvian. Up to now several molecular epidemiology studies has revealed the vast majority of the Brazilian HTLV-1aA strains could be originated from several African lineages [Van Dooren et al., 1998; Yamashita et al., 1999; Alcantara et al., 2003].

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

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

## METHODS

### Study Subjects and Serology

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

### Phylogenetic Analysis

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

The phylogenetic analysis, based on the complete LTR region, included reference strains from different geographic regions and distinct ethnic groups corresponding to all subtypes and described subgroups were selected from the NCBI/Nucleotide Sequence Database (GenBank). To better discuss the epidemiology of HTLV-1 in the Amazon region and to obtain more genetic information about virus

isolates from this region, four sequences (one complete and three partial) generated in a past study [Colin et al., 2003] were also genotyped through the phylogenetic analysis performed in this study. These sequences were also submitted to Genbank after analysis. These sequences were aligned using the Clustal X software [Jeanmougin et al., 1998] and manually edited using the GeneDoc program [Nicholas et al., 1997]. The Tamura Nei (TrN + G) evolutionary model (which takes into account different substitution rates for transversions and transitions, as well as inter-site substitution rate heterogeneity, using a  $\gamma$ -distribution) was selected through the Modeltest software [Possada & Krandall, 1998] as the best model. The Neighbor-Joining (NJ) and Maximum-Likelihood (ML) trees were generated using the PAUP\* 4.0b10 software [Swofford, 1998]. The NJ tree was constructed with an optimized nucleotide substitution rate matrix and a  $\gamma$ -shape parameter (alpha parameter=0.811083). The reliability of the NJ trees was assessed by analyzing 1,000 bootstrap replicates. For the ML tree, an heuristic search was performed with a subtree-pruning-regrafting branch swapping algorithm using the NJ tree as the starting material, including its optimized parameters. The likelihood ratio test (RT) method was used to calculate statistical support for the branches:  $p < 0.001$  (highly significant \*\*) and  $p < 0.005$  (significant \*). Bootstrap and ML supports were added to NJ tree that was drawn with TreeView 1.4 software [Page, 1996]. The genotyping of three partial 3'LTR sequences was performed using the LASP HTLV-1 Automated Genotyping Tool (<http://lasp.cpqgm.fiocruz.br>).

### **Potential Protein Site Analysis**

The genetic information obtained from the proteins were possible through the protein potential sites using the GeneDoc software [Nicholas et al., 1997] and the Prosite tool, as previously described [Mota-Miranda et al., 2007]. The protein sequences included in these analysis were generated in a

previous seroprevalence study [Colin et al., 2003], and all the results were also compared with other Brazilian reference sequences.

### **HTLV-1 Intra-country Diversity**

The genetic distances were measured, within and between four distinct groups formed with LTR strains originated from Brazil (Salvador, Fortaleza, Acre) and Peru. It was used the Tamura Nei model with a distance matrix implemented in the MEGA 3.0 package [Kumar et al., 1994], and the standard error computation was obtained by Bootstrap analysis (1000 replicates).

### **Nucleotide Sequence Accession Numbers**

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



## RESULTS

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

The three *env* gene sequences, corresponding to the gp46 N-terminal region (179-320aa) and the gp21 C-terminal region (1-43aa) were submitted to the GeneDoc software using the Prosite tool to determine potential protein sites. Three different kind of posttranslational modifications were found into the gp46 protein: one PKC phosphorylation site at 310-312 amino acids (aa); one CK2 phosphorylation site at 194-197aa; and three N-glycosylation sites at 222-225aa, 244-247aa and 272-275aa. Otherwise, into the gp21 protein, two different posttranslational modifications were identified: one N-myristylation site at 7-18aa and one PKC phosphorylation site at 22-24a. The potential protein sites described for these three new *env* sequences are the same sites that were already described [Queiroz et al., 2007] for all Brazilian *env* gene sequences previously submitted to GenBank. These confirm the conservative pattern of these proteins and the importance of them to the virus.

To investigate the origin and diversity of HTLV-1 in the Western Amazon region using the mean inter-patient distance, three different groups were formed with the nucleotide sequences: Acre (North Brazil), Fortaleza and Salvador (Northern Brazil). The within groups distance was, 2.7% (Acre), 0.7% (Fortaleza) and 1.2% (Salvador). Comparing the groups, the results demonstrated that the strain

diversity between Acre and Fortaleza was 1.9%, while the divergence between Acre and Salvador was 2.1%, and finally between Fortaleza and Salvador, 1.2%. To better discuss about the origin of HTLV-1 in the Latin American continent, the genetic diversity between Peruvian isolates (2.1%) were also calculated to compare to the genetic distance among Amazon region (Acre) isolates (2.7%).

## DISCUSSION

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

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

The phylogenetic analysis performed in this study has identified two isolates from the Transcontinental (A) subgroup of the Cosmopolitan (a) subtype. This finding is in agreement with others studies that have indicated the wide presence of this subgroup in viral isolates from Brazil [Alcantara et al., 2003]. The contribution of genetic studies indicates at least two hypotheses that give some support to the origin of this subgroup of the HTLV-1 Cosmopolitan Subtype in the New World. The first hypothesis suggests that the virus introduction occurred during the first Paleo-Indian migration across the Bering Straits around 40,000 to 10,000 years ago [Miura et al., 1994]. The second hypothesis suggests that the virus was brought by the post-Columbian African slave trade, around 400 years ago [Van Dooren et al., 1998].

The fact that the HTLV-1 isolates from Rio Branco have grouped into two separate and well-defined clades, and that these two clades demonstrate low divergence, could suggest a recent and separate

spread of the HTLV-1a virus in this Latin American population. A similar conclusion was considered in a recent Peruvian HTLV-1 molecular study [Zehender et al., 2007], when three of ten new HTLV-1 isolates segregated within the main South African cluster, while seven of them clustered into the two already known Latin American clusters.

The higher genetic diversity among Acre strains (2.7%) allow to suppose a previous migration of the virus, across the Bering Strait to the American Continent, when comparing to the lower genetic diversity among Northern Brazil strains (Salvador-1.2%). The last diversity percentage might suggest that the virus was also brought to Brazil, more recently, during the post-Columbian introduction, through the slave trade. The difference between genetic diversity calculated among Acre and Salvador isolates can give support also to the hypothesis that the slaves in Salvador were traded mainly from the West Coast of Africa, while those in the Amazon or North of Brazil came from the East Coast of Africa. Comparing the genetic diversity among Amazon region and Peruvian strains (2.7% x 2.1%), the relative relationship between these isolates could confirm, as postulated before, that the genetic admixture of the black Latin Americans with other Latin American ethnic groups, like Quechuas and Mestizos, giving rise, probably, to HTLV-1a in these indigenous Latin American populations [Van Dooren et al., 1998].

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

the same, as these previously reported in other Brazilian sequences [Queiroz et al., 2007; Mota-Miranda et al., 2007]. The presence of these conserved sites in the Env amino acid sequences might suggest that the virus has maintained them as strategy for escaping from the immune system.

### **ACKNOWLEDGMENT**

This work was partially supported by the Fundação de Amparo a Pesquisa do Estado da Bahia (FAPESB) and PN-DST/AIDS, Ministério da Saúde, Brasil. The authors are grateful to Noilson Lázaro de Souza Gonçalves and Elisabeth Deliege for technical assistance. We thank Dr David Watkins (Visiting Professor, CAPES-FIOCRUZ) for editorial assistance.

**REFERENCES**

- Alcantara LC, de Oliveira T, Gordon M, Pybus O, Mascarenhas RE, Seixas MO, Gonçalves M, Hlela C, Cassol S, Galvão-Castro B 2006 Tracing the origin of Brazilian HTLV-1 as determined by analysis of host and viral genes. *AIDS* 20(5):780.
- Alcantara LCJ, Van Dooren S, Gonçalves MS, Kashima S, Costa MCR, Santos FLN, Bittencourt AL, Dourado I, Andrade Filho A, Covas DT, Vandamme A-M, Galvão-Castro B. 2003. Globin haplotypes of human T-cell lymphotropic virus type I (HTLV-I) infected individuals in Salvador, Bahia, Brazil suggest a Post-columbian African origin of this virus. *J Acquir Immune Defic Syndr* 33: 536-542.
- Bouamir F, Scarlata S, Carter C. 2003. Role of myristylation in HIV-1 gag assembly. *Biochem* 42: 6408-6417.
- Carneiro-Proietti AB, Ribas JG, Catalan-Soares BC, Martins ML, Brito-Melo GE, Martins Filho OA, Pinheiro SR, Araújo AQ, Galvao-Castro B, de Oliveira MS, Guedes AC, Proietti FA. 2002. Infection and disease caused by the human T cell lymphotropic viruses type I and II in Brazil. *Rev Soc Bras Med Trop* 35: 499-508.
- Colin DD, Alcantara LCJ, Santos FLN, Uchôa R, Tavares-Neto J. 2003. Seroprevalence of human T cell lymphotropic virus infection and associated factors of risk in blood donors of Rio Branco city, AC, Brazil (1998-2001). *Rev Soc Bras Med Trop* 36(6):677-683.
- Dourado I, Alcantara JLC, Barreto ML, Teixeira MG, Galvão-Castro B. 2003. HTLV-1 in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. *J Acquir Immue Defic Syndr* 34: 527-531.
- Galvão-Castro B, Loures L, Rodrigues LGM, Sereno A, Ferreira Jr. OC, Franco LGP, Mueller M, Sampaio DA, Santana A, Passos LM, Proietti F. 1997. Geographic 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 The G. 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* 2: 407-410.
- Ishak R, Harrington Jr W, Azevedo VN, Eiraku N, Ishak MOG, Guerreiro JF, Santos SEB, Kubo T, Monken C, Alexander S, Hall WW. 1995. Identification of human T-cell lymphotropic virus type IIa infection in the Kayapó, an Indigenous population of Brazil. *AIDS Res Hum Retrovirol* 11: 813-821.
- Ishak R, Vallinoto ACR, Azevedo VN, Ishak MOG. 2003. The epidemiological aspects of retrovirus (HTLV) infections among Indian populations of the Amazon region of Brazil. *Cad Saúde Pública* 19: 109-118.
- Jeanmougin F, Thompson JD, Goy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with clustalX. *Trends Biochem Sci* 23:403-405.
- Kumar S, Tamura K, Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Cabios* 10:189-191.
- Miura T, Fukunaga T, Igarashi T. 1994. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *PNAS USA* 91: 1124-1127.
- Mota A, Nunes C, Melo A, Romeo M, Boasorte N, Dourado I, Alcantara LC, Galvão-Castro B. 2006. A case-control study of HTLV-infection among blood donors in Salvador, Bahia, Brazil - Associated risk factors and trend towards declining prevalence *Rev Bras Hematol Hemoter* 28(2):120-126.
- Mota-Miranda AC, de-Oliveira T, Moreau DR, Bomfim C, Galvão-Castro B, Alcantara LC 2007. Mapping the molecular characteristics of Brazilian human T-cell lymphotropic virus type 1 Env (gp46) and Pol amino acid sequences for vaccine design. *Mem Inst Oswaldo Cruz* 102(6):741-749.
- Nagy K, Clapham P, Cheingsong-Popov R, Weiss RA. 1983 Human T-cell leukemia virus type I: induction of syncytia and inhibition by patients sera. *Int J Cancer* 32:321-328.



- Nicholas KB, Nicholas HBJ, Deerfield DW. 1997. GeneDoc: Analysis and Visualization of Genetic Variation. *Embnew News* 4:14.
- Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M. 1986. HTLV-I associated myelopathy, a new clinical entity. *Lancet* 1: 1031-1032
- Page RDM. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Cabios Applications Note* 12(4):357-358.
- Palker TJ, Tanner ME, Scearce RM, Streilein RD, Clark ME, Haynes BF. 1989. Mapping of immunogenic regions of human T cell leukemia virus type I (HTLV-I) gp46 and gp21 envelope glycoproteins with env-encoded synthetic peptides and a monoclonal antibody to gp46. *J Immunol* 142: 971–978.
- 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. *PNAS USA* 77:7415–7419.
- Posada D, & Krandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Proietti FA, Carneiro-Proietti ABF, Catalan-Soares BC, Murphy EL. 2005. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* 24:6058-6068.
- Queiroz AT, Mota-Miranda AC, de Oliveira T, Moreau DR, Uripia C de C, Carvalho CM, Galvão-Castro B, Alcântara LC. 2007. Re-mapping the molecular features of the human immunodeficiency virus type 1 and human T-cell lymphotropic virus type 1 Brazilian sequences using a bioinformatics unit established in Salvador, Bahia, Brazil, to give support to the viral epidemiology studies. *Mem Inst Oswaldo Cruz* 102: 133-139.

Reitter JN, Robert EM, Desrosiers RC. 1998. A role for Carbohydrates in immune evasion in AIDS.

Nat Med 4: 679-684.

Swofford DL 1998. PAUP\*: phylogenetic analysis using parsimony (and other methods). Version 4.0b5. Sinauer Associates, Sunderland, Mass.

Vallinoto ACR, Azevedo VN, Santos DEM, Carniceiro S, Mesquita FCL, Hall WW, Ishak MOG, Ishak R. 1998. Serological evidence of HTLV-I and HTLV-II coinfections in HIV-1 positive patients in Belém, state of Pará, Brazil. Mem Inst Oswaldo Cruz 93:407-409.

Van Dooren S, Gotuzo E, Salemi M, Watts D Audenaert E, Duwe S, Ellerbrok H, Grassmann R, Hagelberg E, Desmyter J, Vandamme AM. 1998. Evidence for a post-Columbian introduction of human T-cell lymphotropic virus in Latin America. J Gen Virol 79:2695-2708.

Zehender G, Ebranati E, Bernini F, Maddalena CD, Giambelli C, Collins J, Valverde A, Montin Z, Galli M. 2007. Phylogeny of human T-cell lymphotropic virus type 1 (HTLV-1) in Peru: high degree of evolutionary relatedness with South African isolates. AIDS Res Hum Retrovirol 23(9): 1146-1149.

Yamashita M, Veronesi R, Menna Barreto M, Harrington Jr W J, Sampaio C, Brites C, Badaró R, Andrade-Filho AS, Onkura S, Igarashi T, Takehisa J, Miura T, Chamone D, Bianchini O, Jardim C, Sonoda S, Hayami M. 1999. Molecular epidemiology of human T-cell leukemia virus type I (HTLV-I) in Brazil: The predominant HTLV-Is in South America differ from HTLV-Is of Japan and Africa, as well as those of Japanese immigrants and their relatives in Brazil. Virology 261: 59-69.

Yang YC, Hsu TY, Liu MY, Lin MT, Chen JY, Yang CS. 1997. Molecular subtyping of human T-lymphotropic virus type I (HTLV-I) by a nested polymerase chain reaction-restriction fragment length polymorphism analysis of the envelope gene: two distinct lineages of HTLV-I in Taiwan. J Med Virol 51(1):25-31.



Figure 1- Rooted neighbor-joining tree of Brazilian HTLV-1 LTR region (615bp) isolates. HTLV-1 subtype reference sequences were obtained from the GenBank/EMBL databases. The bootstrap values >50% were considered and added in the tree. Mel5 was used as outgroup. Geographic origin is given between parentheses. The new LTR sequences included in this analysis are in bold. \*\* and \* means that the ML method was highly significant ( $p < 0.001$ ) or significant ( $p < 0.005$ ), respectively.

## 6. DISCUSSÃO

Nossos resultados sugerem que o gene *env*, especialmente a região que compreende o N-terminal de gp46 e o C-terminal de gp21, poderia ser usada para estudos em vacina porque epítomos identificados nesta região apresentam significativa força de ligação às moléculas de HLA e ainda padrões físico-químicos adequados para a produção de anticorpos. Além disso, essa região pode ser alvo de seleção positiva desde que sítios potenciais podem alterar a estrutura terciária do epítomo, e, portanto alterar a resposta imune.

Enquanto o gene *env* deve ser o principal alvo dos anticorpos produzidos naturalmente, o mapeamento molecular do gene *pol* pode ser útil no desenho de protótipos vacinais baseados em mecanismos efetores da resposta imune celular, principalmente CTL, ressaltando a importância de se identificar epítomos específicos, dentro deste gene.

Através das análises computacionais, realizadas neste estudo, foi possível identificar epítomos, CD4 (HLA-DR) e CD8 (HLA-A) específicos, com características conservativas. Acreditamos que a similaridade entre os epítomos identificados, assim como a especificação do alelo de HLA, ao qual se liga o epítomo, pode ser muito útil para direcionar o desenho de vacinas à determinada população com frequência elevada dos alelos importantes, bem como planejar protótipos vacinais que sejam eficientes à maioria dos isolados virais. Além disso, a informação sobre a afinidade de ligação dos epítomos aos alelos de HLA é importante ao passo que o epítomo contendo o sítio de ancoramento selvagem poderia se ligar melhor às moléculas de HLA do que os sítios mutados.

Modificações pós traducionais, em especial, a adição de resíduos, estão diretamente ligadas à mudança conformacional da estrutura protéica e conseqüentemente comprometimento da apresentação adequada do epítomo. Desta forma, o perfil de existência destes domínios em comparação com as

mutações encontradas poderia sugerir haver uma seleção de um ou outro domínio de maneira a dificultar a resposta imune e conseqüentemente favorecer o escape viral.

Estes achados reforçam a importância do gene *env* na resposta imune e no sucesso da infecção viral, e confirmam que esta região é alvo constante da pressão do sistema imune.

A identificação de epítomos a partir de análises computacionais já é utilizada para o desenho de uma vacina para HIV, confirmando a utilidade de ferramentas de bioinformática para selecionar e construir um consenso de epítomos imunogênicos (DE GROOT *et al.*, 2005; FONSECA *et al.*, 2006).

A investigação molecular do envelope viral, especialmente a identificação de mutações em domínios imunodominantes das glicoproteínas do envelope, também pode ser muito útil, no entendimento de fenômenos como a indeterminação no *Western Blot*. Este teste diagnóstico se baseia na especificidade da ligação entre os anticorpos naturais da infecção contra os antígenos virais sintéticos, inclusive a gp46 e a gp21.

Dados oriundos de estudos epidemiológicos, em doadores de sangue, já permitiram identificar taxas de prevalência desta infecção, variando de 0,08% a 2%, nesta população. Além do mais, dados cedidos pela Agência Nacional de Vigilância Sanitária (ANVISA) relatam 6.123 casos de sorologia positiva para o HTLV em doadores de sangue no Brasil, de Janeiro a Outubro de 2001. No entanto, determinadas populações ainda são carentes de estudos de prevalência ainda mais reveladores, como é o caso de populações localizadas em toda a extensão da região amazônica.

A epidemiologia molecular do HTLV-1 permite esclarecer não só a origem, mas também a disseminação deste retrovírus nas diversas populações humanas. Resultados semelhantes, aos encontrados neste estudo, foram relatados por Rego e col. (2008), onde tanto a prevalência da infecção, bem como a epidemiologia molecular do vírus foi avaliada em três distritos populacionais (Taquarandí, Junco e Alegre) do Vale do São Francisco, no estado da Bahia. Os dados da avaliação molecular

reforçam a possibilidade de introdução múltipla do vírus no país, podendo ser tanto de origem Pré como Pós-Colombiana.

Interessantemente, encontramos que uma menor diversidade genética quando comparamos os isolados do Acre, com os isolados de Fortaleza, diferentemente dos achados quando comparado com Salvador. Dados históricos, relatam que durante a segunda guerra mundial, os seringais da Indochina foram tomados pelos japoneses, e o Acre dessa forma representou um grande marco na história Ocidental e Mundial, mudando o curso da guerra a favor dos Aliados e graças aos soldados da borracha oriundos principalmente do sertão do Ceará. Com o alistamento de nordestinos, Getúlio Vargas minimizou o problema da seca do nordeste e ao mesmo tempo deu novo ânimo na colonização da Amazônia. O alistamento compulsório em 1943 era feito pelo Serviço Especial de Mobilização de Trabalhadores para a Amazônia (SEMTA), com sede no nordeste, em Fortaleza, criado pelo então Estado Novo, desta forma, só do nordeste foram para a Amazônia 54 mil trabalhadores, sendo 30 mil deles apenas do Ceará.

Dados epidemiológicos da infecção pelo HTLV demonstram que rotas de transmissão, taxas de prevalência e características genéticas de isolados virais variam em diferentes regiões geográficas. Desta forma, estudos de investigação epidemiológica em diferentes populações e regiões geográficas devem ser motivados com o intuito de se traçar perfis mais detalhados de possíveis rotas migratórias do vírus, dentro do território nacional, além de permitir o conhecimento molecular das cepas virais encontradas em diferentes populações.

## 7. CONCLUSÃO

- Existem, pelo menos, 12 epítomos, no gene *env*, e 4 epítomos no gene *pol*, ligantes específicos de alelos de HLA-I e HLA-II em 12 (*env*) e 4 (*pol*) peptídeos previamente descritos.
- Os epítomos mapeados apresentaram alta similaridade, com variação total de aminoácido entre 9% e 17% para os epítomos do gene *env*, e 1,25% para os epítomos do gene *pol*, sendo um fator importante para o desenvolvimento de protótipos vacinais.
- Entre as mutações identificadas no gene *env*, oito delas são sítios selecionados positivamente por mecanismos de seleção natural, com valor de  $\omega=9,59$  e  $P=1$ .
- A região estudada do gene *env*, referente à segunda metade da gp46 e à primeira metade da gp21, apresenta altos perfis de antigenicidade, hidrofobicidade e acessibilidade, além de assinaturas importantes como sítios de fosforilação para PKC e CK2, sítio de N-glicosilação e miristilação.
- Num corte transversal, realizado em Julho de 2004, foi possível identificar uma soroprevalência de 0,46% do HTLV-1 em doadores de sangue de Rio Branco (Acre).
- Os fragmentos LTR dos isolados gerados e caracterizados (AC181, AC174, AC042, AC069 e AC129), neste estudo, foram classificados como pertencentes ao subgrupo Transcontinental do subtipo Cosmopolita, confirmando a presença majoritária deste subgrupo na população brasileira.
- O perfil de segregação nos grupamentos monofiléticos, bem como os cálculos de diversidade genética permite, ainda, sugerir uma possível introdução Pré-colombiana do HTLV-1, por



migrações pelo Estreito de Bering na população estudada, sem, no entanto, abandonar a possibilidade de introdução Pós-Colombiana.

- As seqüências do gene *env* (AC204, AC057 e AC069), geradas dos isolados virais caracterizados neste estudo, apresentaram assinaturas protéicas, do tipo modificações pós-traducionais, como sítios de fosforilação para PKC e CK2, sítios de N-glicosilação e miristilação.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALCANTARA, LCJ; VAN DOOREN, S; GONÇALVES, MS; KASHIMA, S; COSTA, MC; SANTOS, FL; BITTENCOURT, AL; DOURADO, I; FILHO, AA; COVAS, DT; VANDAMME, AM; GALVÃO-CASTRO, B. Globin haplotypes of human T-cell lymphotropic virus type I-infected individuals in Salvador, Bahia, Brazil, suggest a post-Columbian African origin of this virus. **J Acquir Immune Defic Syndr** 33, 536-542, 2003.

ALCANTARA, LCJ; OLIVEIRA, T; GORDON, M; PYBUS, O; MASCARENHAS, RE; SEIXAS, MO; GONÇALVES, M; HLELA, C; CASSOL, S; GALVÃO-CASTRO, B. Tracing the origin of Brazilian HTLV-1 as determined by analysis of host and viral genes. **AIDS** 20, 780-782, 2006.

BANGHAM, CR. The Immune control and cell-to-cell spread of human T-lymphotropic virus type 1. **J Gen Virol** 84, 3177-3189, 2003.

BANGHAM, CRM & OSAME, M. Cellular immune response to HTLV-1. **Oncogene** 24, 6035–6046, 2005.

BATTINI, JL; MANEL, N; KIM, FJ; KINET, S; TAYLOR, N; SITBON, M. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. **Cell** 115, 449-459, 2003.

BIGGAR, RJ; TAYLOR, ME; NEEL, JV; HJELLE, B; LEVINE, PH; BLACK, FL; SHAW, GM; SHARP, PM; HAHN, BH. Genetic variants of Human T-lymphotropic virus type II in American Indian groups. **Virology** 216, 165-173, 1996.

BITTENCOURT, AL; VIEIRA, MG; BRITES, CR; FARRE, L; BARBOSA, HS. Adult T-Cell Leukemia/Lymphoma in Bahia, Brazil. Analysis of Prognostic Factors in a Group of 70 Patients. **Am J Clin Pathol** 128, 875-882, 2007.

BRADY, J; JEANG, KT; DUVALL, J; KHOURY, G. Identification of p40x-responsive regulatory sequences with the human t cell leukemia virus type 1 long terminal repeat. **J Virol** 61 (7), 2175-2181, 1987.

CARNEIRO-PROIETTI, ABF; RIBAS, JGR; CATALAN-SOARES, BC; MARTINS, ML; BRITOMELO, GEA; MARTINS-FILHO, OA; PINHEIRO, SR; ARAUJO, AQC; GALVAO-CASTRO, B; POMBO DE OLIVEIRA, MS; GUEDES, AC; PROIETTI, FA. Infection and disease caused by the human T cell lymphotropic viruses type I and II in Brazil. **Rev Soc Bras Med Trop** 35(5), 499-508, 2002.

CATALAN-SOARES, B; CARNEIRO-PROIETTI, IAB; PROIETTI FA. Vírus-T linfotrópico humano em familiares de candidatos a doação de sangue soropositivos: disseminação silenciosa. **Rev Panam Salud Publica/Pan Am J Public Health** 16(6), 387-394, 2004.

CATALAN-SOARES, B; CARNEIRO-PROIETTI, ABF; PROIETTI, FA. 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** 21(3), 926-931, 2005.

CAVALLI-SFORZA, LL; MENOZZI, P; PIAZZA, A. **The history and geography of human genes**. Princeton University Press, New Jersey, USA. 1994

CHEN, J; ZEKENG, L; YAMASHITA, M; TAKEISHA, J; MIURA, T; IDO, E; MBOUDJEKA, I; TSAGUE, JM; HAYAMI, M; KAPTUE L. HTLV isolated from a Pygmy in Cameroon is related but distinct from the known Central African type. **AIDS Res Hum Retroviruses** 11, 1529-1531, 1995.

CIMARELLI, A; DUCLOS, CA; GESSAIN, A; CASOLI, C; BERTAZZONI, U. Clonal expansion of human T-cell leukemia virus type II in patients with high proviral load. **Virology** 223, 362-364, 1996.

COLIN, DD; ALCANTARA, LCJ; SANTOS, FLN; UCHOA, R; TAVARES-NETO, J. Prevalência da infecção pelo vírus linfotrópico humano de células T e fatores de risco associados à soropositividade em doadores de sangue da cidade de Rio Branco, AC, Brasil (1998-2001). **Rev Soc Bras Med Trop** 36, 677-683, 2003.

CONSTATINE, NT; CALLAHAN, JD; WATTS, DM. Retroviral Testing. Essential for Quality Control and Laboratory Diagnosis. **CRC Press** 105-107, 1992.

COSKUN, AK & SUTTON, RE. Expression of glucose transporter 1 confers susceptibility to human T-cell leukemia virus envelope-mediated fusion. **J Virol** 79(7), 4150-4158, 2005.

DE GROOT, AS; MARCON, L; BISHOP, EA; RIVERA, D; KUTZLER, M; WEINER, DB; MARTIN, W. HIV vaccine development by computer assisted design: the GAIA vaccine. **Vaccine** 23, 2136-2148, 2005.

DELAMARRE, L; ROSENBERG, AR; PIQUE, C; PHAM, D; CALLEBAUT, I; DOKHÉLAR, MC. The HTLV-1 envelope glycoproteins: structure and functions. **J. Acquir Immune Defic Syndr Hum Retrovirol** 13, Suple. 1, S85-S91, 1996.

DE OLIVEIRA, T; SALEMI, M; GORDON, M; VANDAMME, AM; VAN RENSBURG, EJ; ENGELBRECHT, S; COOVADIA, HM; CASSOL, S. Mapping Sites of Positive Selection and Amino Acid Diversification in the HIV Genome: An Alternative Approach to Vaccine Design? **Genetics** 167, 1047-1058, 2004.

DE THE, G & KAZANJI, M. An HTLV-I/II vaccine: From animal models to clinical trials? **J Acq Immun Def Synd Hum Retrovirol** 13 Suppl 1, S191-S198, 1996.

- DOURADO, I; ALCANTARA, LCJ; BARRETO, ML; DA GLORIA, TM; GALVAO-CASTRO, B. HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics. **J Acquir Immune Defic Syndr** 15; 34(5), 527-531, 2003.
- ELOVAARA, I; KOENIG, S; BREWAH, AY; WOODS, RM; LEHKY, T; JACOBSON, S. High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. **J Exp Med** 177, 1567–1573, 1993.
- FALQUET, L; PAGNI, M; BUCHER, P; HULO, N; SIGRIST, CJA; HOFMANN, K; BAIROCH, A. The Prosite database, its status in 2002. **Nucleic Acids Research** 30, 235-238, 2002.
- FONSECA, SG; COUTINHO-SILVA, A; FONSECA, LAM; SEGURADO, AC; MORAES, SL; RODRIGUES, H; HAMMER, J; KALLÁS, EG; SIDNEY, J; SETTE, A; KALIL, J; CUNHA-NETO, E. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by the HIV-1 infected patients. **AIDS** 20, 2263-2273, 2006.
- GALVÃO-CASTRO, B; LOURES, L; RODRIQUES, LG; SERENO, A; FERREIRA, JOC; FRANCO, LG; MULLER, M; SAMPAIO, AS; PASSOS, LM; PROIETTI, FA. Distribution of human T-lymphotropic virus type I among blood donors: a nationwide Brazilian study [letter]. **Transfusion** 37, 242-243, 1997.
- GESSAIN, A; BARIN, F; VERNANT, JC; GOUT, O; MAURS, L; CALENDER, A; DE THE, G. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. **Lancet** 2, 407-410, 1985.
- GESSAIN, A; GALLO, RC; DE THÉ, G. Genetic variability and molecular epidemiology of Human and Simian T-cell leukemia/ lymphoma virus type 1. **J Acq Imm Def Synd and Hum Retrov** 13 S1, 132-145, 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. **J Acq Immun Def Synd Hum Retrovirol** 13 Suppl 1, S228–S235, 1996.
- GOON, PKC; IGAKURA, T; HANON, E; MOSLEY, AJ; BARFIELD, A; BARNARD, AL; KAFTANTZI, L; TANAKA, Y; TAYLOR, GP; WEBER, JN; BANGHAM, CRM. Human T Cell Lymphotropic Virus Type I (HTLV-I)-Specific CD4<sup>+</sup> T Cells: Immunodominance Hierarchy and Preferential Infection with HTLV-I. **J of Immunol** 172, 735–1743, 2004.
- GREEN, PL & CHEN, ISY. **Fields Virology** (4th edn. Knipe D, Howley P, Griffin D, Lamb R, Martin M and Straus S (eds).) Lippincott Williams & Wilkins: Philadelphia (pp.) 1941–1969, 2001.
- HADLOCK, GK; YANG, Q; ROWE, J; FOUNG, SKH. Epitope Mapping of Human Monoclonal Antibodies Recognizing Conformational Epitopes within HTLV Type 1 gp46, Employing HTLV Type 1/2 envelope Chimeras. **AIDS Res Hum Retroviruses** 18, 57-70, 2002.

- HAHN, BH; SHAW, GM; POPOVIC, MLO, MONICO, A; GALLO, RC; WONG-STAAAL, 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** 34, 613-618, 1984.
- HALL, TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl Acids Symp Ser** 41, 95-98, 1999.
- HANCHARD, B; GIBBS, WN; LOFTERS, W; CAMPBELL, M; WILLIAMS, E; WILLIAMS, N; JAFFE, E; CRANSTON, B; PANCHOOSINGH, LD; LAGRENADE, L; WILKS, R; MURPH, E; BLATTNER, W; MANN, A. **Human Retrovirology: HTLV**. Blattner WA (ed). Raven Press: New York pp. 173-183, 1990.
- HOPP, TP & WOODS, KR. Physico-chemical profiles: A computer program for predicting protein antigenic determinants. **Mol Immunol** 20(4), 483-489, 1983.
- IJICHI, S; IZUMO, S; EIRAKU, N. An autoaggressive process against bystander tissues in HTLV-I-infected individuals: a possible pathomechanism of HAM/TSP. **Medical Hypotheses** 41, 542-547, 1993.
- ISHAK, R; HARRINGTON, JRW; AZEVEDO, VN; EIRAKU, N; ISHAK, MOG; GUERREIRO, JF; SANTOS, SEB; KUBO, T; MONKEN, C; ALEXANDER, S; HALL, WW. Identification of human T-cell lymphotropic virus type IIa infection in the Kayapó, an Indigenous population of Brazil. **AIDS Res Hum Retroviruses** 11, 813-821, 1995.
- ISHAK, R; VALLINOTO, ACR; AZEVEDO, VN; ISHAK, MOG. The epidemiological aspects of retrovirus (HTLV) infections among Indian populations of the Amazon region of Brazil. **Cad Saúde Pública** 19, 901-914, 2003.
- JACOBSON, S. Immunopathogenesis of human T-cell lymphotropic virus type I-associated neurologic disease. **J Infect Dis** 186(2), 187-192, 2002.
- JEANMOUGIN, F; THOMPSON, JD; GOY, M; HIGGINS, DG; GIBSON, TJ. Multiple sequence alignment with clustal X. **Trends Biochem Sci** 23, 403-405, 1998.
- KALYANARAMAN, VS; SARNGADHARAN, MG; ROBERT-GUROFF, M; MIYOSHI, I; GOLDE, D; GALLO, RC. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. **Science**, 218(4572), 571-573, 1982.
- KANNAGI, M; HARADA, S; MARUYAMA, I; INOKO, H; IGARASHI, H; KUWASHIMA, G; SATO, S; MORITA, M; KIDOKORO, M; SUGIMOTO, M; FUNAHASHI, S; OSAME, M; SHIDA, H. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD81 cytotoxic T cells directed against HTLV-I-infected cells. **Int Immunol** 3(8), 761-767, 1991.

KAPLAN, JE; OSAME, M; KUBOTA, H. The risk of development of HTLV-I associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. **J Acquir Immune Defic Syndr** 3,1096-1101, 1990.

KUMAR, S; TAMURA, K; NEI, M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. **Cabios** 10, 189-191, 1994.

LA GRENADE, L. HTLV-I-associated infective dermatitis: past, present, and future. **J Acquir Immune Defic Syndr Hum Retrovirol** 13 Suppl 1, S46-49, 1996.

LAL, RB; BRODINE, S; KAZURA, J; MBIDDE-KATONGA, E; YANAGIHARA, R; ROBERT, C. 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. **J of Clin Microbiol** , 296-299, 1992.

LAL, RB; BUCKNER, C; KHABBAZ, RF; KAPLAN, JE; REYES, G; HADLOCK, K; LIPKA, J; FOUNG, SK; CHAN, L; COLIGAN, JE. Isotypic and IgG subclass restriction of the humoral immune responses to human T-lymphotropic virus type-I. **Clin Immunol Immunopathol** 67(1), 40-49, 1993.

LAIRMORE, MD; JACOBSON, S; GRACIA, F; DE, BK; CASTILLO, L; LARREATEGUI, M; ROBERTS, BD; LEVINE, PH; BLATTNER, WA; KAPLAN, JE. Isolation of human T-cell lymphotropic virus type 2 from Guaymi Indians in Panama. **Proc Natl Acad Sci USA** 87, 8840-8844, 1990.

LIM, DG; BOURCIER, KB; FREEMAN, GJ; HAFLER, DA. Examination of CD81 T Cell Function in Humans Using MHC Class I Tetramers: Similar Cytotoxicity but Variable Proliferation and Cytokine Production Among Different Clonal CD81 T Cells Specific to a Single Viral Epitope. **The J of Immunol** 165, 6214-6220, 2000.

LIU, HF; GOUBAU, P; VAN BRUSSEL, M; VAN LAETHEM, K; CHEN, YC; DESMYTER, J; VANDAMME, AM. The three human T-lymphotropic virus type I subtypes arose from three geographically distinct simian reservoirs. **J Gen Virol** 77, 359-368, 1996.

MAHIEUX, R ; IBRAHIM, F ; MAUCLERE, P ; HERVE, V ; MICHEL, P; TEKAIA, F; CHAPPEY, C; GARIN, B; VAN DER RYST, E; GUILLEMAIN, B; LEDRU, E; DELAPORTE, E; DE THÉ, G; GESSAIN, A. 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** 71, 1317-1333, 1997.

MAHIEUX, R; VCHAPPEY, C; GEORGES-COURBOT, MC; DUBREUIL, G; MAUCLERE, P; GEORGES, A; GESSAIN, A. Simian T-cell lymphotropic virus type I from *Mandrillus sphinx* as a simian counterpart of human T-cell lymphotropic virus type I subtype D. **J Virol** 72, 10316-10322, 1998.

- 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** 69, 1695-1710, 1988.
- MANEL, N; KINET, S; JIM, FJ; TAYLOR, N; SITBON, N; BATTINI, JL. GLUT-1 is the receptor os retrovírus HTLV. **Med Sci** 20 (3), 277-279, 2004.
- MANEL, N; BATTINI, JL; TAYLOR, N; SITBON, N. HTLV-1 tropism and envelope receptor. **Oncogene** 24 (39), 6016-6025, 2005.
- MATSUOKA, M. Human T-cell leukemia virus type I (HTLV-I) infection and the onset of adult T-cell leukemia (ATL). **Retrovirology** 2, 2005.
- MATSUOKA, M & JEANG, KT. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. **Nature** 7, 2007.
- MIURA, T; FUKUNAGA, T; IGARASHI, T; YAMASHITA, M; IDO, E; FUNAHASHI, S; ISHIDAC, T; WASHIO, K; UEDAC, S; HASHIMOTO, K; YOSHIDA, M; OSAME, M; SINGHAL, BS; ZANINOVIC, V; CARTIER, L; SONODA, S; TAJIMA, K; INA, Y; GOJOBORI, T; HAYAMI, M. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. **Proc Natl Acad Sci USA** 91, 1124-1127, 1994.
- MIURA, T; YAMASHITA, M; ZANINOVIC, V; CARTIER, L; TAKEHISA, J; IGARASHI, T; IDO, E; FUJIYOSHI, T; SONODA, S; TAJIMA, K; HAYAMI, M. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. **J Mol Evol** 44, S76-S82, 1997.
- MOCHIZUKI, M; ONO, A; IKEDA, E; HIKITA, N; WATANABE, T; YAMAGUCHI, K; SAGAWA, K; ITO, K. HTLV-I uveitis. **J Acquir Immune Defic Syndr Hum Retrovirol** 13 Suppl 1, S50-56, 1996.
- MORGAN, OS; RODGERS-JOHNSON, P; MORA, C; CHAR, G. HTLV-1 and polymyositis in Jamaica. **Lancet** 18 (2), 1184-1187, 1989.
- MOTA, AC; VAN DOOREN, S; FERNANDES, FM; PEREIRA, SA; QUEIROZ, AT; GALAZZI, VO; VANDAMME, AM; GALVÃO-CASTRO, B; ALCANTARA, LC. The close relationship between South African and Latin American HTLV type 1 strains corroborated in a molecular epidemiological study of the HTLV type 1 isolates from a blood donor cohort. **AIDS Res Hum Retroviruses** 23(4),503-507, 2007.
- MUELLER, N. The epidemiology of HTLV-I infection. **Cancer Causes Control** 2(1), 37-52, 1991.
- MUELLER, N; OKAYAMA, A; STUVER, S; TACHIBANA, N. Findings from the Miyazaki Cohort Study. **J Acq Immun Def Synd Hum Retrovirol** 13 Suppl 1, S2-S7, 1996.

- NAGAI, M; BRENNAN, MB; SAKAI, JA; MORA, CA; JACOBSON, S. CD8+ T-cells are an in vivo reservoir for human T-cell lymphotropic virus type I. **Blood** 98, 1858-1861, 2000.
- NEEL, JV; BIGGAR, RJ; SUKERNIK, RI. Virologic and genetic studies related Amerindian origins to the indigenous people of the Mongolia/ Manchuria/ southeastern Siberia region. **Proc Natl Acad Sci USA** 91, 10737-10741, 1994.
- NICHOLAS, KB; NICHOLAS, HBJ, DEERFIELD, DW. GeneDoc: analysis and visualization of genetic variation. **Embnew News** 4, 14, 1997.
- NISHIOKA, K. HTLV-I arthropathy and Sjogren syndrome. **J Acquir Immune Defic Syndr Hum Retrovirol** 13 Suppl 1, S57-62, 1996.
- OSAME, M; USUKU, K; IZUMO, S; IJICHI, N; AMITANI, H; IGATA, A; MATSUMOTO, M, TARA, M. HTLV-I associated myelopathy, a new clinical entity. **Lancet** 1, 1031-1032, 1986.
- OSAME, M. Pathological Mechanisms of human T-cell lymphotropic virus typeI-associated myelopathy (HAM-TSP). **J Neurovirol** 8 (5), 359-364, 2002.
- PAGE, RDM. TreeView: An application to display phylogenetic trees on personal computers. **Cabios Applications Note** 12(4), 357-358, 1996.
- PAQUETTE, Y; HANNA, Z; SAVARD, P; BROUSSEAU, R; ROBITAILLE, Y; JOLICOUER, P. Retrovirus-induced murine motor neuro disease: mapping the determinant of spongiform degeneration with the envelope gene. **Proc Natl Acad Sci USA** 86 (10), 3896-3900, 1989.
- PARKER, CE; DAENKE, S; NIGHTINGALE, S; BANGHAM, CRM. Activated, HTLV-I-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. **Virology** 188, 628-636, 1992.
- PIQUE, C; CONNAN, F; LEVILAIN, JP; CHOPPIN, J;DOKHÉLAR, MC. Among all Human T-cell Leukemia Virus Type 1 Proteins, Tax, Polymerase, and Envelope Proteins are Predicted as Preferential targets for the HLA-A2 Restricted Cytotoxic T-cell Response. **J of Virol** 4919-4926, 1996.
- 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 USA** 77, 7415-7419, 1980.
- POSADA D, & CRANDALL KA. MODELTEST: testing the model of DNA substitution. **Bioinformatics** 14, 817-818, 1998.
- PROIETTI, FA; CARNEIRO-PROIETTI, ABF; CATALAN-SOARES, B; MURPHY, EL. Global Epidemiology of HTLV-I infection and associated diseases. **Oncogene** 24, 6058-6068, 2005.



- RAMMENSEE, HG; BACHMANN, J; EMMERICH, NPN; BACHOR, OA; STEVANOVIC, S. SYFPEITHI: database for MHC ligands and peptide motifs. **Immunogenetics** 50, 213–219, 1999.
- REGO, FFA; ALCANTARA, LCJ; MOURA NETO, JP; MOTA-MIRANDA, ACA; PEREIRA, OS; GONÇALVES, MS; GALVÃO-CASTRO, B. HTLV-1 molecular study in Brazilian villages with African characteristics giving support to the post-Columbian introduction hypothesis. **AIDS Res and Huma Retrov** 24(5), 673-677, 2008.
- REICH, DE & GOLDSTEIN, DB. Genetic evidence for a paleolithic human population expansion in Africa. **Proc Natl Acad Sci USA** 95, 8119-8123, 1998.
- RODRIGUES, N. **Os africanos no Brasil**. 9: 13-70. Companhia Editora Nacional (5<sup>a</sup> ed), São Paulo, Brasil. 1977
- ROMAN, GC & OSAME, M. Identity of HTLV-I-associated tropical spastic paraparesis and HTLV-I-associated myelopathy. **Lancet** 19;1(8586), 651, 1988.
- ROSENBERG, A; DELAMARRE, L; PREIRA, A; DOKHELAR, MC. Analysis of functional Conservation in the Surface and Transmembrane Glycoprotein Subunits of Human T-Cell Leukemia Virus Type 1 (HTLV-1) and HTLV-2. **J of Virol** 7609–7614, 1998.
- SAGARA, Y; INOUE, Y; SHIRAKI, H; JINNO, A; HOSHINO, H; MAEDA, Y. Identification and Mapping of Functional Domains on Human T-Cell Lymphotropic Virus Type 1 Envelope Proteins by Using Synthetic Peptides. **J of Virol** 70 (3), 1564-1569, 1996.
- SALEMI, M; VAN DOOREN, S; AUDENAERT, E; DELAPORTE, E; GOUBAU, P; DESMYTER, J; VANDAMME, AM. Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have close relatives among African HTLV-I strains. **Virology** 246, 277-287, 1998.
- SALEMI, M. **Molecular investigations of the origin and genetic stability of the human T-cell lymphotropic viruses**. PHD thesis in Katholieke Universiteit Leuven. Rega Instituut. 1999.
- SEGURADO, ACC; BIASUTTI, C; ZEIGLER, R; RODRIGUES, C; DAMAS, CD; JORGE, MSLG; MACHIORI, PE. 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 São Paulo, Brazil. **Mem Inst Oswaldo Cruz** 97, 329-333, 2002.
- 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 USA** 80, 3618-3622, 1983.
- SEIKI, M; EDDY, R; SHOWS, TB; YOSHIDA, M. Nonspecific integration of the HTLV provirus genome into Adult t-cell leukemia cells. **Nature** 309, 640-642, 1984.

SHERMAN, MP; DUBE, S; SPICER, TP; KANE, TD; LOVE, JL; SAKSENA, NK; LANNONE, R; GIBBS, CJ; YANAGIHARA, JR; DUBE, DK; POIESZ, JR. Sequence Analysis of an Immunogenic and Neutralizing Domain of the Human T-Cell Lymphoma/Leukemia Virus Type I gp46 Surface Membrane Protein among Various Primate T-Cell Lymphoma/Leukemia Virus Isolates Including Those from a Patient with Both HTLV-I-associated Myelopathy and Adult T-Cell Leukemia. **Cancer Research** 53, 6067-6073, 1993.

SHIMOYAMA, M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984-87). **Br J Haematol** 79(3), 428-437, 1991.

SHINDO, N; ALCANTARA, LCJ, Van Dooren, S; SALEMI, M; COSTA, MCR; KASHIMA, S; COVAS, DT; TEVA, A; PELLEGRINI, M; BRITO, I; VANDAMME, AM; GALVAO-CASTRO, B. Human Retroviruses (HIV and HTLV) in Brazilian Indians: Seroepidemiological Study and Molecular Epidemiology of HTLV Type 2 Isolates. **AIDS Res and Hum Retrov** 18(1), 71-77, 2002.

SZUREK, PF; YUEN, PH; JERZY, R; WONG, PK. Identification of point mutations in the envelope gene of Moloney murine leukemia virus TB temperature-sensitive paralitogenic mutant ts1: molecular determinants for neurovirulence. **J Virol** 62 (1), 357-360, 1988.

SWOFFORD, DL. **PAUP\*: phylogenetic analysis using parsimony** (and other methods). Version 4.0b5. Sinauer Associates, Sunderland, Mass, 1998

TAYLOR, GP. Pathogenesis and treatment of HTLV-1 associated myelopathy. **Sex Transm Inf** 74, 316-322, 1998.

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

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** 202, 212-223, 1994.

VAN DOOREN, S; GOTUZZO, E; SALEMI, M; WATTS, D; AUDENAERT, E; DUWE, S; ELLERBROK, H; GRASSMANN, R; HAGELBERG, E; DESMYTER, J; VANDAMME, AM. Evidence for a post-Columbian introduction of HTLV-I in Latin America. **J Gen Virol** 79, 2695-2708, 1998.

VERGER, P. **Flux et reflux de la traite des nègres entre le Golfe de Benin et Bahia de Todos os Santos**. Mouton, Paris. 1968

VIANA FILHO, L. **O negro na Bahia**. 31-224. Nova Fronteira (3<sup>a</sup> ed), Rio de Janeiro, Brasil. 1988

WATTEL, E; VARTANIAN, JP; PANNETIER, C; WAIN-HOBSON, S. Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. **J Virol** 69(5), 2863-2868, 1995.

WOLFE, ND; HENEINE, W; CARR, JK; GARCIA, AD; SHANMUGAM, V; TAMOUFE, U; TORIMIRO, JN *et al.*: Emergence of unique primate T-Lymphotropic viruses among central African bushmeat hunters. **Proc Natl Acad Sci USA** 22, 7994-7999, 2005.

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

YAMASHITA, MR; VERONESI, M; MENNA-BARRETO, WJJ; HARRINGTON, C; SAMPIO, C; BRITES, R; BADARO, AS; ANDRADE-FILHO, S; OKHURA, T; IGARASHI, T; MIURA, D; CHAMONE, O; BIANCHINI, C; JARDIM, S; SONODA; HAYAMI, M. Molecular epidemiology of human T-cell leukemia virus type I (HTLV-1) Brazil: the predominant HTLV-1s in South America differ from HTLV-1s of Japan and Africa, as well as those of Japanese immigrants and their relatives in Brazil. **Virology** 261, 59-69, 1999.

YANG, Z. PAML: a program package for phylogenetic analysis by maximum likelihood . **Comput Appl Biosci** 13, 555-556, 1997.

YANG, YC; HSU, TY; LIU, MY; LIN, MT; CHEN, JY; YANG, CS. Molecular subtyping of human T-lymphotropic virus type I (HTLV-I) by a nested polymerase chain reaction-restriction fragment length polymorphism analysis of the envelope gene: two distinct lineages of HTLV-I in Taiwan. **J Med Virol** 51(1), 25-31, 1997.

YANG, Z; NIELSEN, R; GOLDMAN, N; PEDERSEN, AM, Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics** 155, 431-449, 2000.

YASUNAGA, JI & MATSUOKA, M. Human T-Cell Leukemia Virus Type I Induces Adult T-Cell Leukemia: From Clinical Aspects to Molecular Mechanisms. **Cancer Control** 14, 133-140, 2007.

Ficha Catalográfica elaborada pela Biblioteca do  
Centro de Pesquisas Gonçalo Moniz / FIOCRUZ - Salvador - Bahia.

M672e Miranda, Aline Cristina Andrade Mota  
Estudo genotípico de novos isolados do HTLV-1 e caracterização do envelope viral, correlacionando com o desenvolvimento de protótipos vacinais. [manuscrito] / Aline Cristina Andrade Mota Miranda. - 2008.  
106 f. : il. ; 30 cm.

Datilografado (fotocópia).

Mestre (dissertação) – Fundação Oswaldo Cruz. Centro de Pesquisas Gonçalo Moniz, 2008.

Orientador: Prof. Dr. Luiz Carlos Júnior Alcântara, Laboratório Avançado de Saúde Pública.

1. Virus Linfotrópico de Células T Humanas Tipo 1. 2. Genótipos virais.  
3. Vacinas. I.Título.

CDU 616.98:615.371

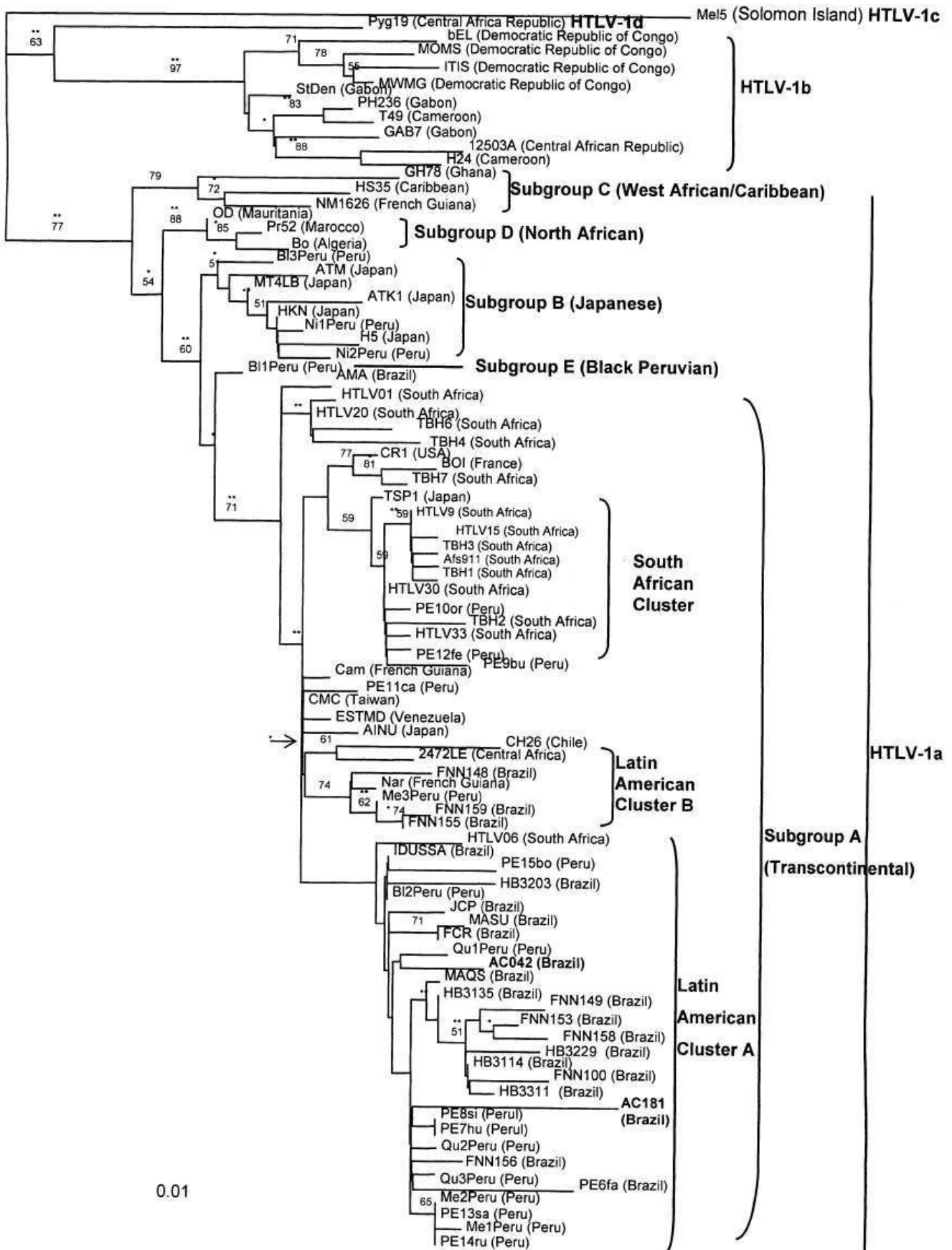


TABLE I

Human T-cell lymphotropic virus type 1 previous published Env (gp46) peptides versus possible human leukocyte antigen (HLA) class I and II epitopes and similarity with Brazilian sequences

HLA class	Peptide sequence	Similarity %	Epitope sequence	SYFPEITHI epitope prediction	Similarity (%)	HLA allele	
I	197-D H I L E P S I P W K S K L L T L V Q L-216 N . . . . . 3/15 <sup>a</sup>	80	195-N L D H I L E P S I P W K-207 . . . N . . . . . 3/15 <sup>b</sup>		80	1/2/3	
	213-L V Q L T L Q S T N Y T C I V C I D R A S L S T-236 . G . . . . . 1/15 <sup>a</sup> . G . I . . . . . A . E . . . . . 1/15 <sup>a</sup> . . . . . A . . . . . 1/15 <sup>a</sup>	80	211-L T L V Q L T L Q S T N Y-223 . . . G . . . . . 2/15 <sup>b</sup> . . . . . I . . . . . 1/15 <sup>b</sup>		80	3/4/5/6/7/8	
	235-S T W H V L Y S P N V S V P S S S S T P-254 . . . . . Y I . . . . . 2/15 <sup>a</sup> . . . . . I . . . . . 5/15 <sup>a</sup>	53	233-S L S T W H V L Y S P N V-245 . . . . . 15/15 <sup>b</sup>		100	▲	
	253-T P L L Y P S L A L P A P H L T L P F N W T H C F D P Q I Q-282 . . . . . 15/15 <sup>a</sup>	100	251-S S T P L L Y P S L A L P-263 . . . . . 15/15 <sup>b</sup>		100	▲	
	277-F D P Q I Q A I V S S P C H N S L-293 . . . . . S . . . . . 1/15 <sup>a</sup>	93	275-C F D P Q I Q A I V S S P-287 . . . . . S . . . . . 15/15 <sup>b</sup>		100	▲	
	288-P C H N S L I L P P F S L S P V P L G S R S R R A V P V A-317 S . . . . . 15/15 <sup>a</sup>	93	286-S S P C H N S L I L P P F-298 . . . S . . . . . 1/15 <sup>b</sup>		93	2/3/7/8/9/10 11/12/13	
	176-L N T E P S Q L P P T A P P L L P H S N L D H I-199 ≈ . . . . . N . . . . . 3/15 <sup>a</sup>	80	179-E P S Q L P P T A P P L-190 ≈ . . . . . N . . . . . 15/15 <sup>b</sup>		100	▲	
	190-L P H S N L D H I L E P S I P W K S K-209 . . . . . N . . . . . 3/15 <sup>a</sup>	80	188-P P L L P H S N L D H I L-200 . . . . . N . . . . . 3/15 <sup>b</sup>		80	1/2/3/4/8/9/10 11/12/13	
	224-Y T C I V C I D R A S L S T W H V L Y S P-244 . . . . . A . . . . . 1/15 <sup>a</sup> . . . . . A . E . . . . . 1/15 <sup>a</sup>	86	222-N Y T C I V C I D R A S L-234 . . . . . A . . . . . 1/15 <sup>b</sup> . . . . . A . E . . . . . 1/15 <sup>b</sup>		85	1/2/3/4/6/8/9/10 11/12/13	
	240-V L Y S P N V S V P S S S S T P L L Y P S L A-262 . . . . . Y I . . . . . 2/15 <sup>a</sup> . . . . . I . . . . . 5/15 <sup>a</sup>	53	238-H V L Y S P N V S V P S S-250 . . . . . Y I . . . . . 1/15 <sup>b</sup> . . . . . I . . . . . 1/15 <sup>b</sup>		85	2/3/4/7/9/10/11 12/13	
	274-W T H C F D P Q I Q A I V S S P C H N S L I L-296 . . . . . S . . . . . 1/15 <sup>a</sup>	93	272-N W T H C F D P Q I Q A I-284 . . . . . S . . . . . 15/15 <sup>b</sup>		100	▲	
	292-N S L I L P P F S L S P V P T L G S R S R R A-314 . . . . . 15/15 <sup>a</sup>	100	290-H N S L I L P P F S L S P-302 . . . . . 15/15 <sup>b</sup>		100	▲	
	Total peptide variation into the sequences: 17%			Total epitope variation into the sequences: 9%			
	II	197-D H I L E P S I P W K S K L L T L V Q L-216 N . . . . . 3/15 <sup>a</sup>	80	195-N L D H I L E P S I P W K S K L L T-213 . . . N . . . . . 3/15 <sup>b</sup>		80	1/2/3/4
		213-L V Q L T L Q S T N Y T C I V C I D R A S L S T-236 . G . . . . . 1/15 <sup>a</sup> . G . I . . . . . A . E . . . . . 1/15 <sup>a</sup> . . . . . A . . . . . 1/15 <sup>a</sup>	80	211-L T L V Q L T L Q S T N Y T C I V C I-229 . . . G . . . . . 1/15 <sup>b</sup> . . . G . I . . . . . A . 1/15 <sup>b</sup> . . . . . A . . . . . 1/15 <sup>b</sup>		80	1/4/5/3
		235-S T W H V L Y S P N V S V P S S S S T P-254 . . . . . Y I . . . . . 2/15 <sup>a</sup> . . . . . I . . . . . 5/15 <sup>a</sup>	53	233-S L S T W H V L Y S P N V S V P S S-251 . . . . . Y I . . . . . 5/15 <sup>b</sup> . . . . . Y I . . . . . 2/15 <sup>b</sup>		53	4/5/3
253-T P L L Y P S L A L P A P H L T L P F N W T H C F D P Q I Q-282 . . . . . 15/15 <sup>a</sup>		100	251-S S T P L L Y P S L A L P A P H L T L-269 . . . . . 15/15 <sup>b</sup>		100	▲▲	
277-F D P Q I Q A I V S S P C H N S L-293 . . . . . S . . . . . 1/15 <sup>a</sup>		93	275-H C F D P Q I Q A I V S S P C H N S L-293 . . . . . S . . . . . 1/15 <sup>b</sup>		93	1/2/5/6	
288-P C H N S L I L P P F S L S P V P L G S R S R R A V P V A-317 S . . . . . 15/15 <sup>a</sup>		93	286-S S P C H N S L I L P P F S L S P V P-304 . . . S . . . . . 1/15 <sup>b</sup>		93	1/2/4/5/6/3	
176-L N T E P S Q L P P T A P P L L P H S N L D H I-199 ≈ . . . . . N . . . . . 3/15 <sup>a</sup>		80	179-E P S Q L P P T A P P L L P H S N L-196 ≈ . . . . . N . . . . . 15/15 <sup>b</sup>		100	▲▲	
190-L P H S N L D H I L E P S I P W K S K-209 . . . . . N . . . . . 3/15 <sup>a</sup>		80	188-P P L L P H S N L D H I L E P S I P W-206 . . . . . N . . . . . 3/15 <sup>b</sup>		80	▲▲	
224-Y T C I V C I D R A S L S T W H V L Y S P-244 . . . . . A . . . . . 1/15 <sup>a</sup> . . . . . A . E . . . . . 1/15 <sup>a</sup>		86	222-N Y T C I V C I D R A S L S T W H V L-240 . . . . . A . . . . . 1/15 <sup>b</sup> . . . . . A . E . . . . . 1/15 <sup>b</sup>		86	1/2/4/5/6/3	
240-V L Y S P N V S V P S S S S T P L L Y P S L A-262 . . . . . Y I . . . . . 2/15 <sup>a</sup> . . . . . I . . . . . 5/15 <sup>a</sup>		53	238-H V L Y S P N V S V P S S S S T P L L-256 . . . . . Y I . . . . . 2/15 <sup>b</sup> . . . . . I . . . . . 5/15 <sup>b</sup>		53	1/2/3/4/5/6	
274-W T H C F D P Q I Q A I V S S P C H N S L I L-296 . . . . . S . . . . . 1/15 <sup>a</sup>		93	272-N W T H C F D P Q I Q A I V S S P C H-290 . . . . . S . . . . . 1/15 <sup>b</sup>		93	1/2/3/4/6	
292-N S L I L P P F S L S P V P T L G S R S R R A-314 . . . . . 15/15 <sup>a</sup>		100	290-H N S L I L P P F S L S P V P T L G S-308 . . . . . 15/15 <sup>b</sup>		100	▲▲	
Total peptide variation into the sequences: 17%			Total epitope variation into the sequences: 17%				

<sup>a</sup>: Peptide mutation frequency in all Brazilian sequences (n=15); <sup>b</sup>: Epitope mutation frequency in all Brazilian sequences (n=15); ≈ Amino acid absent in Brazilian

Sequences; HLA I alleles: 1 (HLA A 1510), 2 (HLA A 5101), 3 (HLA A 0201), 4 (HLA A 03), 5 (HLA A 01), 6 (HLA A 6801), 7 (HLA A 026), 8 (HLA A2705), 9 (HLA A 2402), 10 (HLA A 2709), 11 (HLA A 4402), 12 (HLA A 0702), 13 (HLA A 08) and 14 (HLA A 0203); HLA II alleles: 1 (HLA DRB1 0101), 2 (HLA DRB1 1101), 3 (HLA DRB1 0701), 4 (HLA DRB1 0301), 5 (HLA DRB1 1501) and 6 (HLA DRB1 0401); ▲ represent all HLA-I and ▲▲ HLA-II alleles, and the underlined sites represent the positive selected codons.