

**Michelle Camargo**

Caracterização genética e das pressões seletivas atuantes na região do envelope e das variantes GWGR do subtipo B brasileiro do Vírus da Imunodeficiência Humana Tipo 1

Dissertação apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina para obtenção do Título de Mestre em Ciências.

**São Paulo**

**2009**

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**Michelle Camargo**

Caracterização genética e das pressões seletivas atuantes na região do envelope e das variantes GWGR do subtipo B brasileiro do Vírus da Imunodeficiência Humana Tipo 1

Dissertação apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina para obtenção do Título de Mestre em Ciências.

Orientador: Prof. Dr. Luiz Mário Ramos Janini  
Co-orientador: Prof. Dr. Ricardo Sobhie Diaz

**São Paulo**

**2009**

Camargo, Michelle.

**Caracterização genética e das pressões seletivas atuantes na região do envelope e das variantes GWGR do subtipo B brasileiro do Vírus da Imunodeficiência Humana Tipo 1.** / Michelle Camargo -- São Paulo, 2009.

xvi, 85f.

Dissertação (Mestrado) – Disciplina de Doenças Infecciosas e Parasitárias da Universidade Federal de São Paulo –UNIFESP/EPM.

Título em inglês: Genetic characterization and selective pressure of HIV-1 *env* region of Brazilian samples harboring GWGR motif at the tip of the V3 loop.

1. HIV-1 2. Envelope 3. GWGR 4. Filogenia

"Vocês jovens doutores e cientistas do futuro, não se deixem macular por um ceticismo estéril e nem sejam desencorajados pela tristeza que as nações passam em certas horas. Não fiquem bravos com seus adversários. Nenhuma teoria científica já foi aceita sem oposição. Vivam na paz de suas bibliotecas e laboratórios. Perguntem-se antes: "O que fiz pela minha instrução?" E ao avançarem gradualmente: "O que estou conseguindo?" Até que chegue a hora que terão a imensa felicidade de pensar que contribuíram de algum modo para o bem-estar e progresso da humanidade".

(Louis Pasteur - "A história de Louis Pasteur" 1936).

Apoio financeiro da FAPESP

Processo nº 06/53818-7

Projeto nº 06/52988-6

Aos meus pais, Lourdes e Laurindo.  
Tudo que sou hoje devo a vocês  
e isto os tornam tão vitoriosos quanto eu.  
Amo vocês!

## AGRADECIMENTOS

A Deus, “graças te dou, visto que por modo assombrosamente maravilhoso me formaste; as Tuas obras são admiráveis e a minh’alma o sabe muito bem” Salmos 139:14.

Aos meus pais, Laurindo e Lourdes, pela força e paciência, pelo carinho e apoio para que eu chegasse até aqui.

Ao Profº Dr. Luiz Mário Ramos Janini, meu orientador, pelo precioso auxílio, por todas as oportunidades cedidas e por sua admirável busca pelo progresso científico; pela força e incentivo necessários, por acolher e amar seus alunos como um pai ama seus filhos.

Ao Profº Dr. Ricardo Sobhie Diaz, diretor do Laboratório de Retrovirologia – UNIFESP meu co-orientador, pela confiança depositada, pelo imenso espírito científico e pela intrínseca capacidade estimuladora.

À Profª Dra. Maria Cecília Araripe Sucupira, supervisora do Laboratório de Retrovirologia - UNIFESP, pela amizade e atenção, pelas oportunidades oferecidas, pela confiança depositada, pelo apoio, ajuda e carinho indescritivelmente cedidos e para sempre lembrados; meu porto seguro.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo apoio financeiro do processo número 06/53818-7 e projeto número 06/52988-6, sem os quais seria impossível a realização desta pesquisa.

Aos amigos que passaram durante este tempo e aos permanecem no laboratório de Retrovirologia – UNIFESP: Adriana, Alessandra, Alexandre, Andréa, Ana Carolina, Camila F., Camila S., Carla, Celina, Cristiano, Daniel, Daniela, Edsel, Élcio, Elizabeth, Érika, Giana, Jean, Juliana G., Juliana M., Lucas, Mara, Maria Clara, Mariana, Márcia, Michel, Michelle Z., Patrícia, Rafael, Rodrigo C., Rodrigo M., Sandra Mara, Shirley, Thais, Wagner e demais, que compartilharam alegrias e transformaram tristezas em aprendizado sabendo cultivar uma amizade que o tempo amadureceu.

Ao secretário Charlys, pela colaboração e divertida espontaneidade, pela ajuda sempre presente.

Ao aluno de mestrado e amigo Leonardo Zambotti Villela pela atenção e ajuda; por ter sido uma luz no fim do túnel.



A todos os meus demais amigos, tão queridos e importantes para mim, não mencionados aqui para que por um imperdoável descuido eu deixe de citar alguém, mas evidentes em meu coração pela amizade e pelo carinho demonstrados, pelos risos partilhados, pela força e apoio doados; amigos cujo nome segue gravado em meu coração, não sendo nem o tempo capaz de apagar.

Muito, muito obrigada!

## SUMÁRIO

Epígrafe .....	iv
Dedicatória .....	vi
Apoio Financeiro .....	v
Agradecimentos .....	vii
Lista de figuras .....	xii
Lista de tabelas .....	xiv
Resumo .....	xv
Abstract .....	xvi
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
<b>1.1. História e Descoberta do HIV .....</b>	<b>1</b>
<b>1.2. Epidemiologia .....</b>	<b>2</b>
<b>1.3. O HIV-1: Organização Genômica .....</b>	<b>4</b>
<b>1.3.1. Genes Estruturais .....</b>	<b>6</b>
<b>1.3.1.1. Gene <i>gag</i> .....</b>	<b>6</b>
<b>1.3.1.2. Gene <i>pol</i> .....</b>	<b>7</b>
<b>1.3.1.3. Gene <i>env</i> .....</b>	<b>8</b>
<b>1.3.2. Genes Regulatórios .....</b>	<b>9</b>
<b>1.3.2.1. Gene <i>tat</i> .....</b>	<b>9</b>
<b>1.3.2.2. Gene <i>rev</i> .....</b>	<b>9</b>
<b>1.3.2.3. Gene <i>nef</i> .....</b>	<b>10</b>
<b>1.3.3. Genes Acessórios .....</b>	<b>10</b>
<b>1.3.3.1. Gene <i>vif</i> .....</b>	<b>10</b>
<b>1.3.3.2. Gene <i>vpu</i> .....</b>	<b>11</b>
<b>1.3.3.3. Gene <i>vpr</i> .....</b>	<b>11</b>
<b>1.4. Ciclo Replicativo do HIV-1 .....</b>	<b>11</b>
<b>1.5. Diversidade Viral .....</b>	<b>13</b>
<b>1.6. Formas Recombinantes Circulantes (CRFs) .....</b>	<b>14</b>
<b>1.7. Diversidade Genética do HIV-1 .....</b>	<b>15</b>
<b>1.8. Envelope .....</b>	<b>17</b>
<b>1.8.1. GPGR <i>versus</i> GWGR .....</b>	<b>17</b>
<b>2.0. OBJETIVOS .....</b>	<b>21</b>
<b>2.1. Objetivo Principal .....</b>	<b>21</b>
<b>2.2. Objetivos Secundários .....</b>	<b>21</b>

<b>3.0. CASUÍSTICA</b> .....	22
<b>3.1. Aspectos Éticos</b> .....	22
<b>3.2. Critérios de Inclusão no Estudo</b> .....	22
<b>3.3. Seleção das Amostras</b> .....	22
<b>4.0. METODOLOGIA</b> .....	24
<b>4.1. Extração do Ácido Nucléico do HIV-1</b> .....	24
<b>4.2. Reação em Cadeia pela Polimerase (PCR) em duas etapas (<i>nested</i>) para amplificação do DNA proviral</b> .....	24
<b>4.2.1. Genes <i>env</i></b> .....	25
<b>4.2.2. Genes <i>pol</i></b> .....	26
<b>4.2.3. Genes <i>gag</i></b> .....	28
<b>4.3. Detecção dos Produtos Amplificados</b> .....	29
<b>4.4. Purificação dos Produtos Amplificados</b> .....	29
<b>4.5. Reação de Seqüenciamento</b> .....	31
<b>4.6. Precipitação da Reação de Seqüenciamento</b> .....	33
<b>4.7. Análise do Seqüenciamento</b> .....	33
<b>4.8. Análise dos Subtipos e Detecção de Recombinação</b> .....	35
<b>4.9. Análise Filogenética das Seqüências Obtidas</b> .....	35
<b>4.9.1. Compilação e Alinhamento das Seqüências</b> .....	35
<b>4.9.2. Inferência Filogenética</b> .....	36
<b>4.9.3. Métodos de Detecção de Seleção</b> .....	38
<b>4.9.3.1. Determinação de dN e dS pelo método de distância par-a-par</b> .....	38
<b>4.9.3.2. Determinação de dN e dS pelo método de verossimilhança máxima</b> .....	40
<b>5.0. RESULTADOS</b> .....	43
<b>5.1. Extração</b> .....	43
<b>5.2. Extringência da Reação</b> .....	44
<b>5.2.1. Gene <i>env</i></b> .....	44
<b>5.2.2. Gene <i>pol</i></b> .....	53
<b>5.2.3. Gene <i>gag</i></b> .....	54
<b>5.3. Análise dos Subtipos e Detecção de Recombinação</b> .....	56
<b>5.4. Análise da Pressão Seletiva</b> .....	65
<b>6.0. DISCUSSÃO</b> .....	73
<b>7.0. CONCLUSÕES</b> .....	78
<b>8.0. REFERÊNCIAS</b> .....	79

## LISTA DE FIGURAS

<b>Figura 1.</b> Distribuição Global de Adultos Infectados pelo HIV-1 em 2005 .....	3
<b>Figura 2.</b> Partícula do HIV-1 .....	5
<b>Figura 3.</b> Organização Genômica do HIV-1 .....	6
<b>Figura 4.</b> Alça da Região V3 do HIV-1 .....	9
<b>Figura 5.</b> Ciclo Replicativo do HIV-1 .....	13
<b>Figura 6.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (bandas inespecíficas) .....	44
<b>Figura 7.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (teste de concentração de magnésio, dNTP e iniciadores) .....	45
<b>Figura 8.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (variação da temperatura de anelamento) .....	46
<b>Figura 9.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (novos iniciadores) .....	48
<b>Figura 10.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (novas bandas inespecíficas) .....	50
<b>Figura 11.</b> Eletroforese em Gel de Agarose – Gene <i>env</i> (final) .....	51
<b>Figura 12.</b> Eletroforese em Gel de Agarose – Gene <i>pol</i> .....	53
<b>Figura 13.</b> Eletroforese em Gel de Agarose – gene <i>gag</i> .....	54
<b>Figura 14.</b> Exemplo de Amostra Subtipo B, Segundo a Metodologia do SIMPLOT .....	58
<b>Figura 15.</b> Exemplo de Amostra Subtipo B, Segundo a Metodologia do SIMPLOT .....	58
<b>Figura 16.</b> Exemplo de Amostra Subtipo não B, Segundo a Metodologia do SIMPLOT .....	59
<b>Figura 17.</b> Exemplo de Amostra Subtipo não B, Segundo a Metodologia do SIMPLOT .....	59
<b>Figura 18.</b> Árvore Construída por <i>Maximum Likelihood</i> para o Gene <i>env</i> .....	60
<b>Figura 19.</b> Árvore Construída por <i>Maximum Likelihood</i> para o Gene <i>pol</i> .....	61
<b>Figura 20.</b> Árvore Construída por <i>Maximum Likelihood</i> para o Gene <i>gag</i> .....	62
<b>Figura 21.</b> Árvore Construída por <i>Maximum Likelihood</i> para as Amostras GWGR e GPGR com os Genes <i>env</i> , <i>pol</i> e <i>gag</i> .....	63
<b>Figura 22.</b> Média do Comportamento das Mutações Sinônimas e Não Sinônimas em Toda a Região Codificada .....	66

## LISTA DE TABELAS

<b>Tabela 1.</b> Descrição dos iniciadores externos utilizados na primeira etapa da <i>nested</i> -PCR do gene <i>env</i> .....	25
<b>Tabela 2.</b> Descrição dos iniciadores externos utilizados na segunda etapa da <i>nested</i> -PCR do gene <i>env</i> .....	26
<b>Tabela 3.</b> Descrição dos iniciadores externos utilizados na primeira etapa da <i>nested</i> -PCR do gene <i>pol</i> .....	26
<b>Tabela 4.</b> Descrição dos iniciadores externos utilizados na segunda etapa da <i>nested</i> -PCR do gene <i>pol</i> .....	27
<b>Tabela 5.</b> Descrição dos iniciadores externos utilizados na primeira etapa da <i>nested</i> -PCR do gene <i>gag</i> .....	28
<b>Tabela 6.</b> Descrição dos iniciadores externos utilizados na segunda etapa da <i>nested</i> -PCR do gene <i>gag</i> .....	28
<b>Tabela 7.</b> Descrição dos iniciadores utilizados na reação de seqüenciamento para os genes <i>env</i> , <i>pol</i> e <i>gag</i> .....	32
<b>Tabela 8.</b> Código IUPAC .....	34
<b>Tabela 9.</b> Quantidade de amostras extraídas, amplificadas e seqüenciadas para GWGR .....	43
<b>Tabela 10.</b> Quantidade de amostras extraídas, amplificadas e seqüenciadas para GPGR .....	42
<b>Tabela 11.</b> Número de amostras GWGR subtipadas e utilizadas no estudo .....	56
<b>Tabela 12.</b> Número de amostras GPGR subtipadas e utilizadas no estudo .....	56
<b>Tabela 13.</b> Identificação das amostras e seus subtipos .....	57
<b>Tabela 14.</b> Valores de substituições sinônimas e não sinônimas por <i>Nei &amp; Gojobori</i> .....	67
<b>Tabela 15.</b> Valores de Diversidade Genética e Desvio Padrão .....	67
<b>Tabela 16.</b> Análise por <i>Maximum Likelihood</i> dos genes <i>env</i> , <i>pol</i> e <i>gag</i> para as amostras GWGR e GPGR .....	68

## RESUMO

Apesar do subtipo B ser o mais predominante na epidemia de aids no Brasil, existem evidências que esse subtipo não é homogêneo. O subtipo B apresenta no Brasil uma assinatura no tetrâmero da coroa V3 da gp120 que é a presença de um triptofano (W) ao invés de prolina (P) que é o aminoácido mais freqüente nessa posição do genoma do HIV-1. Aproximadamente metade dos isolados do subtipo B no Brasil possuem um triptofano (W) no tetrâmero do *loop* V3 da gp120 (GWGR). Essa freqüência não é observada em outros subtipos e nem mesmo no subtipo B em outras epidemias de aids no mundo. Para explorar mais as características das variantes GWGR e GPGR foi realizada uma análise detalhada incluindo inferências filogenéticas, análise de diversidade genética e de pressão seletiva em seqüências dos genes *env*, *pol* e *gag*. Os resultados da análise de diversidade genética e pressão seletiva mostraram que os valores da razão entre dN/dS entre os genes são distintos. Para uma aferição mais precisa da pressão seletiva sugerida pelo método par-a-par foram realizadas análises usando um método de verossimilhança que estima os valores de dN/dS em cada códon individualmente presente no alinhamento de seqüências. Os resultados obtidos com essa análise indicaram códons sob seleção positiva ao longo da seqüência da região dos genes *env*, *pol* e *gag* do HIV-1 em ambas as variantes. Para realmente definir se amostras GWGR e GPGR são variantes distintas ou amostras que contêm apenas uma assinatura diferenciada em *env*, foi composta uma seqüência concatêmera e comparada a distribuição de ambas as amostras em um agrupamento filogenético. Essa análise mostrou que as amostras GWGR não são variantes distintas do subtipo B brasileiro, mas apenas seqüências com assinaturas diferentes. A distribuição encontrada demonstrou que não houve separação das amostras GWGR e GPGR e sim uma total mistura entre elas. Essas análises mostraram que os vírus que contêm W ou P no *loop* V3 não estão evoluindo diferencialmente na epidemia de aids no Brasil e não são variantes distintas. Possivelmente, os vírus que contêm a variante GWGR entraram cedo na epidemia (efeito fundador), sofrendo disseminação no Brasil.

## ABSTRACT

Despite the subtype B being the most predominant in the epidemic of aids in Brazil, there is evidence that this subtype is not homogeneous. The subtype B in Brazil presents a signature tetramer in the crown of the V3 that gp120 is the presence of a tryptophan (W) instead of proline (P) which is the most common amino acid position in the genome of HIV-1. Approximately half of the isolates of subtype B in Brazil have a tryptophan (W) tetramer in the V3 loop of gp120 (GWGR). This frequency is not observed in other subtypes and even in other subtype B in the aids epidemics in the world. To further explore the characteristics of variants and GWGR GPGR, was performed a detailed analysis including phylogenetic inferences, analysis of genetic diversity and selective pressure in sequences of genes *env*, *gag* and *pol*. Analysis of genetic diversity and selective pressure showed that the values of the ratio of dN / dS between the genes are different. For a more precise measurement of selective pressure method suggested by pair-a-pair analysis was performed using a method of estimating the likelihood values of dN / dS in each individual codon in the alignment of sequences. The results obtained with this analysis indicated codons under positive selection along the region's sequence of genes *env*, *gag* and *pol* of HIV-1 in both variants. To really define whether samples GWGR and GPGR are distinct variations or samples that contain only a signature in deference *env*, was composed a string concatâmera and compared the distribution of both samples in a phylogenetic grouping. This analysis showed that the samples are not GWGR distinct variants of subtype B Brazilian, but only sequences with different signatures. The distribution has found that there was no separation of the samples and GWGR GPGR, but a total mixture between them. These tests showed that the virus that contains W or P in the V3 loop are not differentially evolved in the aids epidemic in Brazil and are not distinct variants. Possibly the virus containing the variant GWGR came early in the epidemic (founder effect), suffering spread in Brazil.

## 1 – INTRODUÇÃO

### 1.1 – HISTÓRIA E DESCOBERTA DO HIV

A aids surgiu no início da década de 80 em várias regiões do mundo caracterizada por anormalidades imunológicas, como queda significativa no número de linfócitos (células CD4<sup>+</sup>), acompanhada de infecções oportunistas, síndromes neurológicas e tumores, porém ainda não era conhecido o agente causador dessa síndrome. A primeira indicação de que a aids era causada por um retrovírus veio em 1983, quando Barré-Sinoussi e colaboradores, reconheceram a atividade da enzima transcriptase reversa característica dos retrovírus nas amostras estudadas (Barré-Sinoussi *et al.*, 1983). Em 1986 foi proposto que os retrovírus associados a aids fossem denominados de HIV (*Human Immunodeficiency Viruses*) devido à similaridades estruturais e biológicas existentes entre os membros do grupo (Barré-Sinoussi *et al.*, 1983).

Estudos epidemiológicos nos anos de 1981 e 1982 identificaram o contato sexual e a exposição ao sangue contaminado como as principais vias de transmissão da aids (Jafee *et al.*, 1983). A síndrome foi descrita inicialmente em usuários de drogas injetáveis e em homens homossexuais e bissexuais, mas posteriormente foi reconhecida a transmissão do vírus também pela atividade heterossexual. O comportamento de risco sexual no início da epidemia causou o rápido aumento da transmissão do HIV ocasionando uma pandemia (McCusker *et al.*, 1998). Com a detecção da existência da transmissão materno-fetal (Ammann *et al.*, 1983; Oleske *et al.*, 1983) conseguiu-se identificar as três principais vias de transmissão do HIV: sangue, contato sexual e transmissão vertical.

São conhecidos atualmente dois tipos de HIV: HIV-1 e HIV-2, ambos pertencentes à família *Retroviridae*. Esses dois tipos antigênicos do HIV possuem origens zoonóticas distintas, o HIV-1 e o HIV-2 derivam do SIVcpz e do SIVagm, respectivamente. Tanto o HIV-1 quanto o HIV-2, dividem-se em diferentes subtipos que se distinguem por suas organizações genômicas e relações filogenéticas. O HIV-1 possui o gene *vpu* e está dividido em três grupos: grupo M (*main ou major*), composto por nove subtipos nomeados: A-D, F-H, J e K (as variantes dos subtipos A e F são ainda segregadas como sub-subtipos A1 ou A2 e F1 ou F2, respectivamente); grupo O (*outlier*) e N (*new*). Além desses, 43 formas recombinantes (CRF) circulam na epidemia. O HIV-2 está dividido em sete subtipos: A, B, C, D, E, F e G, e possui tanto o gene *vpu*, também presente no HIV-1, quanto o gene *vpx* (Robertson, Anderson *et al.*, 2000).



Em 1986, a primeira droga “anti-HIV” foi aprovada pelo FDA (*Food and Drug Administration*), órgão americano que regulamenta e autoriza a comercialização de medicamentos nos EUA e, posteriormente, em todo mundo. A droga foi designada de AZT e também por zidovudina. O AZT (zidovudina), desenvolvido em 1964, era utilizado, inicialmente, como quimioterápico, mas devido ao alto grau de toxicidade para os pacientes nas doses utilizadas, sua prescrição fora abandonada. No início da década de 70, experimentos em laboratório mostraram que o AZT possuía também atividade antiviral e em 1985 foi descrita sua atividade “anti-HIV” (Colson *et al.*, 1991).

Novas drogas foram desenvolvidas e, em 1992, foi apresentado o primeiro estudo envolvendo terapia combinada a partir da utilização concomitante de duas drogas. Em 1996, foi desenvolvida uma nova terapia que combina a utilização concomitante de três ou mais drogas, a denominada Terapia Anti-retroviral Altamente Ativa – HAART (do inglês: *Highly Active Anti-retroviral Therapy*). Porém, o veloz aparecimento de vírus resistentes às diferentes combinações destes fármacos, impeliu o desenvolvimento de novos fármacos e novas estratégias terapêuticas (De Clerc, 1998).

Hoje, a pandemia de aids constitui um grave problema de saúde pública, mas aquela doença antes conhecida como terminal mudou de perfil e passou a ser encarada como uma doença crônica, com tratamento disponível, dando aos indivíduos infectados pelo HIV-1 esperança de vida.

O grande desafio atualmente é fazer com que o tratamento possa ser disponibilizado a todas as pessoas infectadas no mundo, não apenas em países ricos ou com programas que beneficiam o portador, como é o caso do Brasil.

## **1.2 – EPIDEMIOLOGIA**

Entre as enfermidades infecciosas emergentes, a aids destaca-se pela grande magnitude e extensão dos danos causados às populações, de modo que as características da infecção pelo HIV e suas repercussões têm sido exaustivamente discutidas pela comunidade científica e pela sociedade em geral (Brito *et al.*, 2000).

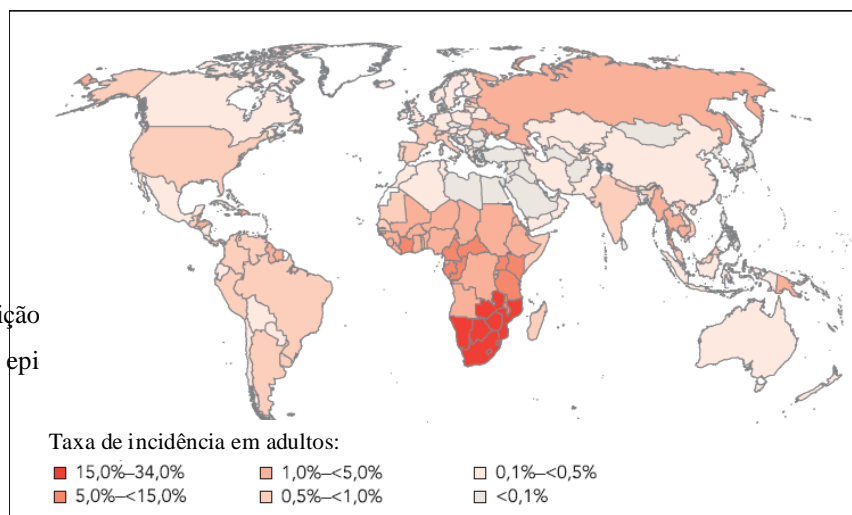
Segundo dados divulgados em 2006, pelo panorama geral da epidemia da aids da UNAids (*Joint United Nations Programme* 2004) divisão da Organização Mundial de Saúde (OMS), até

dezembro de 2005, cerca de 38,6 milhões (33,4 – 46,0 milhões) de pessoas encontravam-se infectadas pelo HIV em todo o mundo (**figura 1**). Estima-se que ocorrem aproximadamente 4,1 milhões (3,4 – 6,2 milhões) de novas infecções pelo HIV e que cerca de 2,8 milhões de pessoas morrem por causa da aids.

Os primeiros casos de aids no Brasil foram identificados em indivíduos da região Sudeste, em 1982. Ao longo dos anos, a aids avançou rapidamente em nosso país e, segundo o Boletim Epidemiológico da Coordenação Nacional de DST/AIDS, programa do Ministério da Saúde, o número oficial de casos notificados de aids no Brasil, até junho de 2006, foi de 433.067 pessoas. Contudo, observa-se a redução proporcional dos casos notificados de transmissão vertical. Ainda, segundo o boletim, o número de óbitos em decorrência da aids no período de 1980 a 2005 foi de 183.074 mortes. No ano de 2005, foi registrado no Sistema de Informações sobre Mortalidade, um total de 11.026 óbitos por aids, representando taxa de mortalidade de 6,0/100.000 habitantes, variando de acordo com a região de residência.

A taxa de mortalidade de homens e mulheres aponta para um quadro de estabilidade nos últimos anos. Essa estabilização nos índices de mortalidade deve-se, em grande parte, ao tratamento anti-retroviral disponibilizado aos portadores do HIV/aids em nosso país. Porém, o índice de novos casos ainda é crescente no Brasil e em todo o mundo.

**Figura 1.** Distribuição  
(Panorama geral da epi



em 2005

### 1.3 – O HIV-1: ORGANIZAÇÃO GENÔMICA

O HIV, semelhante a outros retrovírus, é composto de um envoltório fosfolipoprotéico (envelope) contendo uma cápsula protéica (capsídeo) e por um complexo ribonucleoprotéico (nucleocapsídeo), onde encontra-se o material genômico do vírus. Possui aproximadamente 100nm de diâmetro.

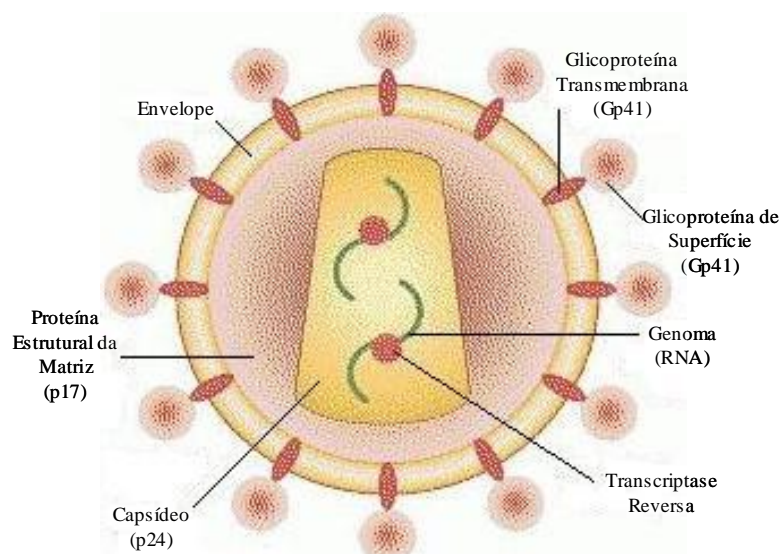
O envelope viral consiste em uma membrana fosfolipídica derivada da célula do hospedeiro e de um complexo protéico formado pelas glicoproteínas gp120 e gp41. A gp120 contém os sítios de ligação aos receptores celulares e o principal domínio antigênico, a alça da região V3. A gp41 é a responsável pela entrada do vírus na célula hospedeira através da fusão das membranas. A proteína precursora do envelope, gp160 (*env*), é direcionada ao retículo endoplasmático da célula hospedeira sendo encaminhada à via secretora celular onde será processada em gp120 e gp41 pelas enzimas celulares e transportada para a superfície celular.

Abaixo dessa camada lipídica existe uma matriz protéica composta pela proteína p17, que envolve o capsídeo. O capsídeo, localizado no interior do envelope, é composto pela proteína p24. Associadas internamente ao *core* há as proteínas *vif* e *nef* e do lado externo a proteína *vpr*.

No interior do capsídeo encontra-se o nucleocapsídeo composto pelo material genômico viral, duas moléculas de RNA (ácido ribonucléico) de polaridade positiva, as proteínas p1, p2, p6, p7 e por três enzimas virais: transcriptase reversa (RT) (DNA <ácido desoxirribonucléico> polimerase RNA dependente) com a subunidade RNase H (p 51/p 66), protease (p 10) e integrase (p 11) (Coffin, 1996).

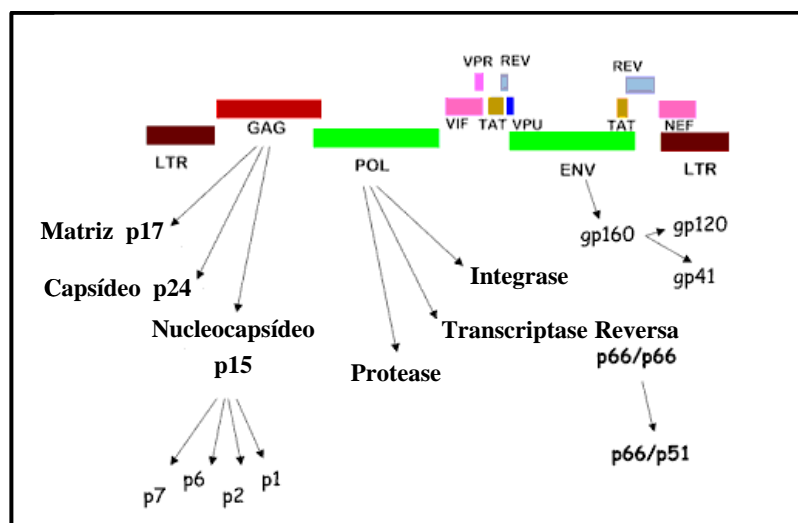
O RNA viral de aproximadamente 9,8 Kb, com várias fases de leitura (ORF), é limitado em suas extremidades por dois Terminais Longos de Repetição (LTR), constituído por genes estruturais (*gag*, *pol* e *env*), genes regulatórios (*tat*, *rev* e *nef*) e genes acessórios (*vpr*, *vpu* e *vif*).

A função do LTR é iniciar a transcrição viral. O transcrito primário do HIV-1 é um mRNA (ácido ribonucléico mensageiro) único, representando todo o genoma viral, que será processado pelo mecanismo de corte e junção (*splicing*) durante a fase inicial do ciclo replicativo formando três proteínas virais: *tat*, *rev* e *nef*.



**Figura 2.** Partícula do HIV-1. As setas indicam as estruturas da partícula viral do HIV-1.

*Tat* (p14) é responsável pela transativação transcricional; interage com o LTR aumentando a transcrição. O acúmulo de *rev* (p19) determina a mudança da fase de replicação, ocorrendo as sínteses das proteínas estruturais *gag*, *pol* e *env* e das acessórias *vpr*, *vpu* e *vif*. *Nef* (p27) atua diretamente na fase inicial da replicação, inibindo a expressão dos receptores CD4 impedindo a entrada de novos vírus na célula; controla a expressão do MHC, diminui a expressão de IL-2 e parece exercer o papel de supressor da expressão do LTR. *Vpr* (p18) auxilia na replicação viral, *vpu* (p16) estimula a liberação dos vírions e *vif* (p23) regula a maturação viral aumentando a infectividade e transmissão para outras células; auxilia também na síntese de cDNA (DNA complementar) proviral (Coffin, 1996).



**Figura 3.** Organização genômica do HIV-1. As setas indicam, de maneira esquemática, as proteínas codificadas pelos genes do HIV-1.

### 1.3.1 – GENES ESTRUTURAIS

#### 1.3.1.1 – GENE *gag*

É um componente da partícula viral que constitui o capsídeo. O gene codifica uma proteína precursora de 9,2 Kb, a p55, que sofre *splicing* a partir da ORF (*Open Reading Frame*), localizada na porção inicial do gene *pol*, e é clivada pela protease viral, dando origem a quatro proteínas que constituem o cerne do HIV: p17 (proteína da matriz); p24 (principal proteína do capsídeo viral); p6 (proteína rica em prolina, envolvida na maturação da partícula viral e que auxilia na inclusão da proteína acessória, *vpr* nos vírions) e p7 (proteína de ligação com o ácido nucléico). As proteínas do gene *gag* executam diversas funções principais durante a montagem viral, incluindo: 1. formação do esqueleto estrutural do vírion; 2. encapsidação do genoma viral; 3. taxiondo a partícula nascente para exportação da célula e 4. aquisição da bicamada lipídica e glicoprotéica associadas ao envelope. Esses processos requerem que as proteínas interajam em processos proteína-proteína, proteína-RNA e proteína-membrana (Coffin, 1996).

Alguns estudos sugerem que a presença de resposta citotóxica direcionada aos diferentes epítomos do gene *gag* esteja associada a uma progressão lenta para a aids (Wellensiek *et al.*, 2006).

### 1.3.1.2 – GENE *pol*

Região gênica, cuja sobreposição se dá no início do gene *pol* e término do gene *gag*. Codifica três enzimas funcionais necessárias à replicação viral: 1. protease: responsável pela formação do gene *gag* que ocorre pelo processo de autocatálise do *gag/pol*; 2. transcriptase reversa: possui duas atividades catalíticas que incluem atividade de DNA polimerase, responsável pela polimerização do DNA a partir do RNA e a atividade da ribonuclease H (Rnase H), que degrada o componente RNA das moléculas híbridas RNA-DNA que foram formadas; 3. integrase: responsável pela integração do genoma viral ao genoma da célula hospedeira.

A protease (p10) é produzida através de uma clivagem autocatalítica da proteína *gag/pol* 160 que ocorre em *cis* (intramolecular) ou em *trans* (intermolecular) durante o processo de morfogênese do vírion (Coffin, 1996). Mutações sítio específicas demonstraram que a protease é uma enzima importante para a replicação já que partículas não infectantes contendo *gag/pol* não clivadas são produzidas quando essa enzima é inativada por uma mutação (Coffin, 1992). Possui 99 aminoácidos de comprimento com peso molecular 10K. Através de análises de cristalografia, mutações sítio-específicas e especificidade de substrato foi comprovado que sua forma funcional é um homodímero. Cada monômero é formado pela duplicação de quatro elementos estruturais: um grampo (que contém a primeira alça), uma grande alça, a qual contém o sítio catalítico, uma  $\alpha$  hélice e um segundo grampo (*flaps*), que se move para permitir a entrada e saída de substrato ou inibidor, de até 7 aminoácidos de comprimento. A protease é uma enzima relativamente conservada, principalmente em locais importantes para a manutenção de sua estrutura e função (Hoffman, 2002).

A transcriptase reversa (p51/p66) é produzida através de duas clivagens da proteína *gag/pol* pela protease. A primeira clivagem dá origem a uma seqüência de peptídeos de 66 kDa (p66) que forma um homodímero, que então é novamente clivado perto da carboxila terminal dando origem agora a um heterodímero formado pela p66 e p51. As duas subunidades possuem a mesma seqüência de peptídeos, embora sua estrutura e função sejam diferentes. A subunidade p66 possui tanto atividade de polimerase como RNase H, enquanto a subunidade p51 possui apenas atividade de polimerase. Essas duas subunidades se interagem assimetricamente para formar uma única *polymerase cleft* que se liga a um *primer-template*, a um dNTP, a um inibidor não competitivo e a um tRNA (RNA transportador). As duas subunidades possuem os mesmos subdomínios, porém são orientados de forma diferente, o que resulta em uma p51 mais compacta, incapaz de se ligar isoladamente ao ácido nucléico, e uma p66 mais estendida. Os três subdomínios que formam o sítio ativo da polimerase são chamados de “dedos”, “polegar” e

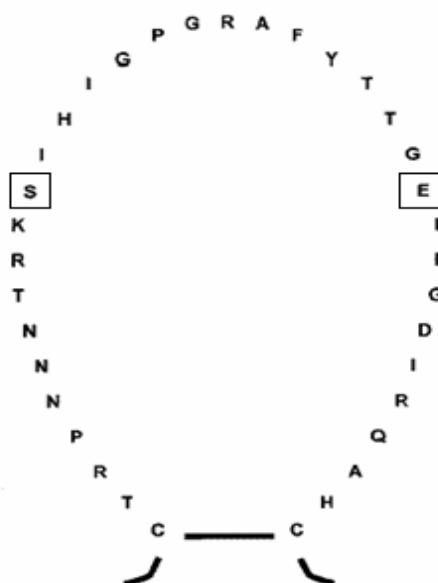
“palma” por analogia dessa estrutura à mão direita. O quarto subdomínio é chamado de “conexão”, pois se situa entre o sítio da polimerase e o sítio ativo da RNase H na p66. Como os subdomínios de conexão são orientados identicamente, o diferente arranjo das interações são feitos pelos “dedos”, “polegar” e “palma” em cada uma das duas subunidades (Coffin, 1992).

### 1.3.1.3 – GENE *env*

Codifica uma proteína precursora, a gp160, que ao ser clivada origina duas glicoproteínas: gp120 e gp41. A gp120 é uma proteína hidrofílica encontrada na superfície externa da membrana plasmática de células infectadas altamente glicosilada. Possui 24 sítios de glicosilação ligados ao N, sendo 13 conservados em vários isolados virais e 17 modificados por cadeias laterais de carboidratos que influenciarão na função e nas propriedades imunológicas da proteína madura (Doms *et al.*, 1993). Possui 18 resíduos de cisteína altamente conservados que formam 9 pontes de dissulfeto definindo várias regiões funcionais e influenciando na estrutura da proteína. Formada por 5 domínios variáveis (V1, V2, V3, V4 e V5) intercalados por cinco domínios conservados (C1, C2, C3, C4 e C5).

O domínio variável V3 é formado por 35 aminoácidos arranjados dentro de uma alça formada pela ponte de dissulfeto entre as cisteínas 296 e 330 (numerados segundo De Jong *et al.*, 1992). Esse domínio tem um importante papel em várias propriedades virais como tropismo celular, citopatogenicidade e fusogenicidade. Deleções nesse domínio podem afetar a infectividade viral através do bloqueio de ligação com o CD4.

A seqüência GPGR (aminoácidos 310, 311, 312 e 313) localizado no topo (ou coroa) da alça é um domínio conservado, enquanto as seqüências próximas às cisteínas, na base da alça, são relativamente variadas. Embora seja um domínio conservado, variantes da seqüência GPGR (característico do subtipo B) já foram descritas.



**Figura 4 – Alça da região V3 do *env***  
Em destaque os aminoácidos 306 e 320, respectivamente (Fouchier *et al*, 1995).

### 1.3.2 – GENES REGULATÓRIOS

#### 1.3.2.1 – GENE *tat*

Codifica a proteína de regulação *tat* (do inglês: *trans-acting transcription transactivator*), que é a principal reguladora transcricional do LTR e é essencial para a replicação viral *in vitro* e *in vivo*. A *tat* é uma proteína nuclear do HIV-1 que contém 101 resíduos de aminoácidos, codificada por dois éxons. O segundo éxon, extremamente conservado entre os isolados de HIV-1 e o primeiro, imbuído das funções de ativação transcricional. Por meio da região transativadora responsiva (TAR), localizada a montante do sítio de iniciação de transcrição do RNA, a *tat* aumenta a processividade da RNA polimerase II. (Sheridan *et al.*, 1991).

#### 1.3.2.2 – GENE *rev*

Codifica a proteína de regulação com 116 resíduos aminoácidos denominada *rev*, necessária para a replicação viral. Essa proteína atua após a transcrição e promove o transporte das moléculas de RNAm das proteínas estruturais e acessórias do núcleo celular ao citoplasma para que ocorra a tradução; controla a proporção entre RNAm, processado e não processado, através da ligação nos elementos de resposta à *rev*, presente em todos os transcritos virais que ainda não



sofreram *splicing* e é responsável pela regulação da síntese de proteínas virais, isto é, equilibra a proporção de produção de proteínas reguladoras em relação às estruturais (Goff, 2003; Petelin & Trono, 2003).

### **1.3.2.3 – GENE *nef***

Codifica a proteína *nef*, que contém aproximadamente 29 kDa, expressa durante todo o ciclo de vida do HIV na célula infectada, concomitantemente a *tat* e *rev*. Ademais, outra característica importante, é a produção de anticorpos e CTL contra esse gene, pois os níveis de ambos também são altos durante toda a infecção no paciente.

A proteína *nef* está envolvida em diversos mecanismos utilizados na manipulação da célula infectada: para auxiliar na replicação viral, no escape das pressões imunológicas, na patogênese e expansão viral. Além disso, a *nef* interage com diversos fatores celulares (Roeth & Collins, 2006). A *nef* interage também com diversas proteínas celulares que, em parte, se relacionam com sinalização intracelular. Estudos estruturais dessa proteína identificaram alguns domínios envolvidos na interação proteína-proteína e favorece um modelo onde *nef* age como um adaptador para acumular complexos de sinalização na célula infectada. Além disso, vírus que apresentam deleção no gene *nef* são menos virulentos e estão associados com a progressão lenta para aids. Essas características promovem a associação de *nef* à patogenicidade do HIV (Das & Jameel, 2005).

## **1.3.3 – GENES ACESSÓRIOS**

Os genes acessórios produzem proteínas que podem alterar drasticamente o curso e a severidade da infecção viral, replicação e progressão da doença.

### **1.3.3.1 – GENE *vif***

O gene *vif* (fator de infectividade viral), codifica uma proteína de 23 KDa (*vif*), que é produzida tardiamente durante a replicação viral. Além disso, esse gene interage com a proteína celular APOBEC-3G, impedindo que a mesma seja encapsulada junto com as novas partículas virais e ocasione destruição de parte do genoma viral via hipermutação (Goff, 2003).

### 1.3.3.2 – GENE *vpu*

Codifica uma proteína de 16KDa (*vpu*). Essa proteína aumenta a taxa de liberação do vírus na célula infectada, facilitando a maturação e montagem de novos vírus e também auxilia na degradação das moléculas de CD4 recém-sintetizadas, antes que estas sejam transportadas para a membrana citoplasmática (Bour & Strebel, 2003).

### 1.3.3.3 – GENE *vpr*

Codifica a proteína *vpr* de 15KDa, traduzida a partir de um RNAm que sofreu um único evento de *splicing*. É encapsulada com o vírion, participando no transporte do complexo de pré-integração do genoma viral para o núcleo da célula infectada. Participa no processo de parada do ciclo celular na fase G2 do ciclo, beneficiando o vírus por garantir maior taxa de expressão protéica. Está envolvida nos processos de apoptose, localização nuclear do complexo de pré-integração, ativação do LTR e de outros promotores heterólogos (Tungaturthi *et al.*, 2003).

## 1.4 – CICLO REPLICATIVO DO HIV-1

A entrada do HIV na célula hospedeira se dá via interação entre a proteína de superfície gp120 com o receptor celular CD4 e com co-receptores celulares auxiliares, como: CCR-5, CCR-2 e CXCR-4 (Levy, 1996). Os receptores tipo CD4 são glicoproteínas pertencentes à superfamília das imunoglobulinas com aproximadamente 55-KDa. Esses receptores possuem domínios extracelulares que servem de sítios de ligação com alta afinidade para a gp120 do HIV-1 (Deng, Liu *et al.*, 1996).

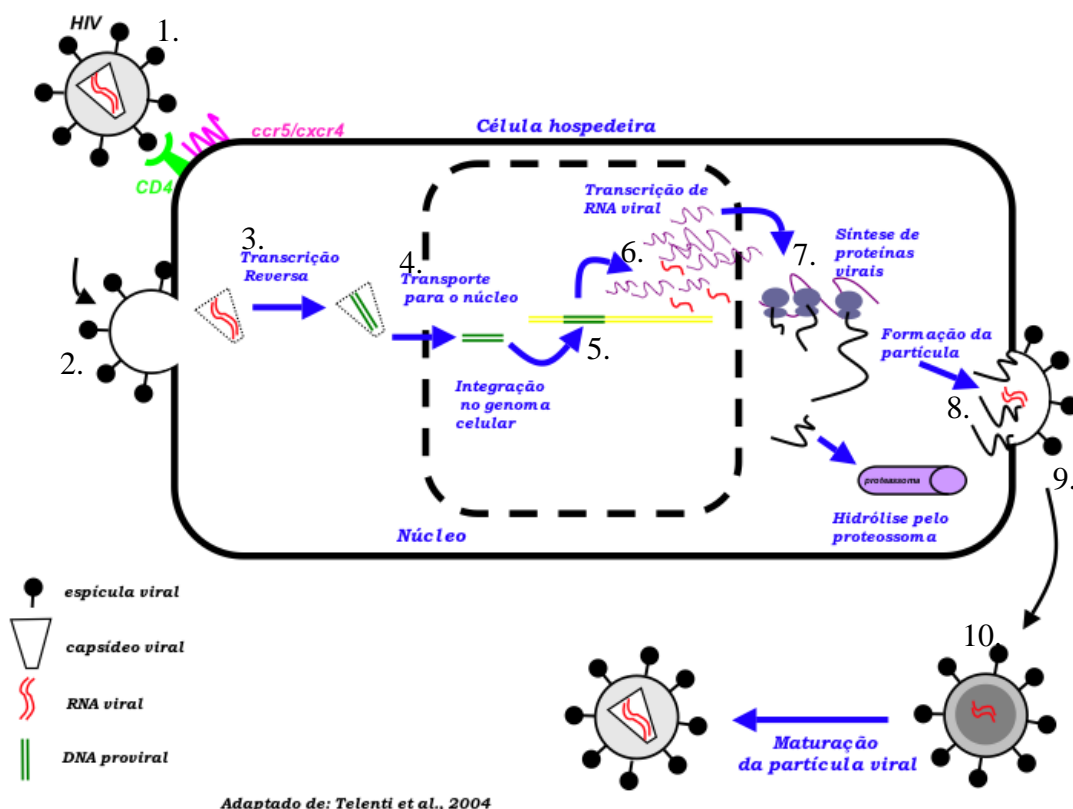
A ligação do CD4 com a gp120 induz mudanças conformacionais que expõem uma região específica dessa glicoproteína (região V3). Essa região passa a interagir com os co-receptores celulares. O resultado dessa interação é a exposição da gp41 que possui o peptídeo de fusão responsável pela interação entre o envelope viral com a membrana celular (Clapham & Weiss, 1997). Após a fusão o nucleocapsídeo é liberado e começa simultaneamente a transcrição reversa e a translocação do restante da partícula viral do citoplasma para o núcleo da célula hospedeira. Sob a ação da transcriptase reversa é produzido um DNA complementar (cDNA) a partir das duas fitas de RNA viral. Uma vez no núcleo, o cDNA genômico viral é integrado ao genoma celular via ação da enzima integrase, constituindo o provírus. O provírus pode permanecer inativo, tornando a

infecção pelo HIV latente, ou pode ser expresso produzindo novas partículas virais. Integrado no genoma da célula hospedeira o genoma do HIV possui cerca de 9,8 Kb, que constituem os genes estruturais *gag* (*group-specific antigen*), *pol* (*polymerase*) e *env* (*envelope*); os genes regulatórios *tat* (*trans-activator*), *rev* (*regulator of expression of virion proteins*), e *nef* (*negative factor*), e os genes acessórios *vif* (*virion infectivity factor*), *vpr* (*viral protein r*), *vpu* (*viral protein u*) e *vpx* (*viral protein x*), sendo esse último somente encontrado no HIV-2. Os terminais 5' e 3' do DNA proviral formam longas seqüências idênticas repetidas denominadas LTR (*long terminal repeat*), que abrigam os elementos regulatórios transcricionais do vírus, sendo que o seu terminal 5' serve como local de acoplamento do complexo transcricional celular (Coffin, 1996).

A descrição dos genes e das proteínas virais é essencial para o entendimento do ciclo replicativo do HIV. O gene *gag* codifica a poliproteína p55, que é precursora da proteína de matriz (p17), da proteína do capsídeo (p24) e das proteínas do nucleocapsídeo (p1, p2, p6 e p7). Devido à sobreposição, em parte, dos genes *gag* e *pol*, eles codificam a proteína (p160) que é clivada gerando as enzimas protease (p10), transcriptase reversa (p66) e integrase (11). A enzima transcriptase reversa (p66) pode ser dividida em duas porções p51 e p15 (RNaseH), essa última responsável pela degradação da fita molde de RNA. O gene *env* codifica a poliproteína gp160 que é clivada por proteases celulares do hospedeiro formando as glicoproteínas do envelope viral (gp120/gp41) (Levy 1993).

A expressão gênica do HIV-1 é dividida em duas fases: fase regulatória precoce e fase estrutural tardia. Na fase precoce são produzidas as proteínas *tat*, *rev* e *nef* e na fase tardia, além das proteínas estruturais, as proteínas *vif*, *vpr*, *vpu* (Cullen & Greene, 1990). A proteína *tat* é um potente transativador viral necessário para a replicação do HIV *in vivo* e *in vitro*. Tem sido proposto que *tat* atua como ativador do promotor viral, a proteína *rev* atua regulando a expressão do mRNA viral e *nef* atua como modulador de sinais intracelulares (Kestler, Ringler *et al.*, 1991).

Os genes que codificam as proteínas acessórias são importantes para a replicação viral. Tem sido demonstrado que a infectividade do HIV é reduzida cerca de 1000 vezes na ausência de *vif*, devido a sua participação na síntese do DNA viral e na ativação da transcriptase reversa. As proteínas codificadas pelo gene *vpr* são importantes para a translocação do complexo viral do citoplasma para o núcleo da célula hospedeira, e as proteínas *vpu* facilitam o brotamento das partículas virais (Strebel, Daugherty *et al.*, 1987).



8.) formação da partícula; 9.) montagem e liberação da partícula viral; 10.) maturação da partícula viral .

## 1.5 – DIVERSIDADE VIRAL

A recombinação é um dos mecanismos mais importantes para a geração de diversidade genética do HIV e a sua frequência está estimada em dois a três eventos por genoma proviral gerado em cada ciclo replicativo (Jetzt, Yu *et al.*, 2000). O mecanismo de recombinação do HIV-1 é um fenômeno complexo e ainda não inteiramente elucidado.

A recombinação ocorre principalmente durante a transcrição reversa. Em um determinado momento dessa transcrição a enzima viral transcriptase reversa pode saltar de uma das fitas de RNA (fita doadora) para a outra fita de RNA (receptora), gerando um DNA mosaico originado das duas moléculas de RNA. Quando esse salto ocorre devido a uma descontinuidade na fita de RNA doadora esse evento é chamado de recombinação por transferência forçada de cadeia. Essa descontinuidade leva a pausas forçadas da retrotranscrição na cadeia doadora e a sua continuação na molécula de RNA receptora (Robertson *et al.*, 1995; Jetzt *et al.*, 2000; Negroni & Buc 2001). O salto da enzima de uma fita de RNA para a outra pode ocorrer também quando existe algum

obstáculo na fita de RNA doadora, por exemplo, quando esse RNA possui estruturas secundárias. Esse mecanismo é chamado de recombinação por transferência de cadeia independente de pausas na retrotranscrição, pois facilitaria a transferência de cadeia sem necessidade de pausas na enzima viral ou quebras na seqüência do RNA doador (Robertson *et al.*, 1995; Jetzt *et al.*, 2000; Negroni & Buc 2001).

O evento de recombinação vem propiciando o aparecimento de genomas mosaicos entre as variantes do HIV-1. Isso ocorre quando uma célula é duplamente infectada com dois vírus pertencentes a variantes distintas. Nesse ciclo replicativo viral existe a possibilidade de serem produzidos três tipos de descendência. Dois tipos poderão ser homozigóticos, tendo duas cadeias de RNA provenientes dos vírus parentais e um terceiro tipo de vírus heterozigótico, tendo duas cadeias diferentes de RNA, uma proveniente de cada vírus parental. A infecção de uma nova célula por esse vírus heterozigoto gera o mosaico entre duas cadeias diferentes de RNA, pois durante a retrotranscrição, a enzima transcriptase reversa salta de uma cadeia de RNA para outra produzindo o DNA complementar, que é um mosaico constituído por genes e fragmentos de cada um dos RNAs parentais (Robertson *et al.*, 1995; Jetzt *et al.*, 2000; Negroni & Buc 2001).

## **1.6 – FORMAS RECOMBINANTES CIRCULANTES (CRFs)**

Acredita-se que pelo menos 10% dos isolados do HIV-1 em circulação sejam recombinantes (McCutchan, 2000). Os recombinantes que se disseminam o suficiente para estabelecer novos focos epidêmicos são chamados de Formas Recombinantes Circulantes (CRF). As características que permitem definir uma CRF são: ter sido detectada em pelo menos três indivíduos não relacionados epidemiologicamente, pelo menos três genomas desse recombinante devem ter sido seqüenciados integralmente, e eles devem possuir uma estrutura molecular e composição de variantes homogênea (Robertson *et al.*, 2000).

Atualmente são descritas 43 CRFs, que apresentam diferenças nas estruturas moleculares, na composição dos subtipos e nas prevalências e distribuições geográficas (Thomson *et al.*, 2002). As CRFs são designadas por números que identificam a ordem de descoberta e pelas letras correspondentes aos subtipos que a constituem. Se forem constituídas por mais de dois subtipos são consideradas complexas, sendo designadas de CRFcp (Robertson, *et al.*, 2000).

A caracterização de amostras recombinantes do HIV é de vital importância para o direcionamento das medidas de detecção e intervenção sobre as formas que estão prevalecendo na

população. Essa caracterização permitiria um aprimoramento da capacidade dos testes diagnósticos de detectarem as formas recombinantes do HIV-1 (Thomson *et al.*, 2002b).

A caracterização das formas recombinantes poderia também contribuir para um eventual desenvolvimento de estratégias de vacinas contra o HIV, pois um dos fatores responsáveis pela inabilidade para o desenvolvimento de vacinas é a dificuldade de se elucidar a resposta imune protetora do hospedeiro contra os diversos mutantes do HIV (Thomson *et al.*, 2002b). Essa caracterização proporcionaria dados adicionais sobre a diversidade e características genéticas de um provável vírus mais adaptado, contribuindo para a geração de métodos de imunização mais abrangentes sobre as diversas formas virais.

### 1.7 – DIVERSIDADE GENÉTICA DO HIV-1

Quase-espécie é uma mistura heterogênea de vírus com genomas relacionados, posto que a população se originou de um ancestral comum, onde cada genoma viral difere do outro em decorrência da evolução desse vírus no hospedeiro (Diaz *et al.*, 1997). As características essenciais das quase-espécies, satisfeitas por HIV, são devidas ao enorme contingente populacional e as taxas elevadas de mutação (Nowak, 1991).

Os mecanismos que geram essa diversidade incluem: mutação, hipermutação, inserção, deleção e recombinação intergênica (Preston *et al.*, 1988; Diaz *et al.*, 1995). Tem sido demonstrada a ocorrência de mutações do HIV-1 *in vivo* e *in vitro*, sendo provenientes da sua alta taxa de replicação e pela elevada frequência de erro de incorporação de nucleotídeo da transcriptase reversa na sua atividade de polimerização. Essa frequência de erro é estimada em cerca de 1-10 incorporações por genoma do HIV em cada ciclo replicativo (Preston *et al.*, 1988).

A hipermutação é definida como um processo em que o cDNA adquire várias mutações em um ciclo replicativo. Preponderantemente, é observada a substituição de G (guanina) por A (adenina) (Janini *et al.*, 2001). Inserção e deleção envolvem normalmente mudanças em múltiplos de três nucleotídeos, levando a inserção ou deleção de um ou mais aminoácidos na proteína viral. Essas alterações têm se fixado predominantemente na região do envelope viral, que é considerada a região mais variável do HIV (Vartanian *et al.*, 1991).

Análises filogenéticas das seqüências virais revelaram que os isolados do HIV-1 são pertencentes a três grupos: M (*main*), O (*outlier*) e N (*non-M non-O*) (Simon *et al.*, 1998; Robertson *et al.*, 2000). Devido a sua variabilidade genética são descritos nove subtipos diferentes do HIV-1 pertencentes ao grupo M, sendo esse grupo o de maior prevalência no mundo. Esses

subtipos são subdivididos de A-D, F-H, J e K. Os subtipos F e A podem ainda ser subdivididos nos sub-subtipos F1 e F2 e A1 e A2 (Robertson *et al.*, 2000). A distribuição geográfica e o padrão de disseminação dos diferentes subtipos do HIV-1 são bastante heterogêneos. O subtipo B é o principal responsável pela disseminação da aids fora da África (<http://www.unaids.org>). Entretanto, o subtipo C é o mais freqüente sendo o principal responsável causador da epidemia na Índia e na África (Shankarappa *et al.*, 1999). Estima-se que aproximadamente 39,5 milhões de pessoas estão infectadas pelo HIV no mundo e que a maioria dos casos (aproximadamente 24,7 milhões de pessoas) concentram-se nos países da região subsaariana. (<http://www.unaids.org>). Notavelmente, na África, ocorrem todos os subtipos já descritos, além de inúmeras formas recombinantes. No Brasil, estima-se que aproximadamente 600 mil pessoas estejam infectadas pelo HIV-1, sendo que existem aproximadamente 240 mil casos registrados. Somente no estado de São Paulo estão localizados 108 mil dos casos (<http://www.aids.gov.br>). No Brasil existe uma alta freqüência de três subtipos do HIV-1: B, F, e C sendo que o subtipo B é o de maior prevalência (cerca de 75% das infecções), F (aproximadamente 10%) e o subtipo C com uma prevalência maior nos estados do sul, predominantemente Rio Grande do Sul, com cerca de 30% das infecções (Morgado *et al.*, 1994; Sabino *et al.*, 1994; Sabino *et al.*, 1996; Rossini *et al.*, 2001; Tanuri *et al.*, 1999; Soares *et al.*, 2003; Brindeiro *et al.*, 2003), além das CRFs.

Proporções tão altas de infecção por CRFs são observadas na África (<http://www.bioafrica.net>) e no sul da Ásia (McCutchan *et al.*, 2000). No Brasil, a CRF\_BC representa 11% dos vírus circulantes na região sul do país (Brindeiro *et al.*, 2003; Soares *et al.*, 2003). Entretanto, outras formas recombinantes circulantes têm sido descritas no Brasil como, CRF28\_BF e CRF29\_BF, as quais são formadas exclusivamente pelos subtipos B e F. Podem também desempenhar um importante papel na epidemia brasileira e na da América do Sul (Sanabani *et al.*, 2006; De Sá Filho *et al.*, 2006). Ademais, foi relatada a primeira ocorrência de um recombinante africano, CRF\_02/AG, no Rio de Janeiro (Couto-Fernandez *et al.*, 2005).

O acúmulo da diversidade genética do HIV em um determinado indivíduo resulta principalmente da taxa de erros da transcriptase reversa, que é cerca de  $1,7 \times 10^{-4}$  erros por base incorporada por ciclo de replicação, com uma taxa de fixação do erro de aproximadamente  $2,5 \times 10^{-5}$  (Mansky *et al.*, 1995) e pelo alto índice replicativo do vírus, que é da ordem de  $10^{10}$  novos vírus por dia (Ho *et al.*, 1995).

Estudos realizados principalmente com indivíduos sintomáticos revelaram que os vírus mudam para escapar da resposta do sistema imune do hospedeiro (McKeating *et al.*, 1989). O aumento da diversidade genética durante o curso da infecção pode também estar relacionado à progressão da doença. O aumento no número de vírus variantes pode ser um fator importante

levando à falência do sistema imune. Teoricamente, o controle da replicação do HIV-1 é perdido quando a diversidade genética do vírus dentro do indivíduo infectado excede a capacidade do sistema imune do mesmo para responder de forma eficaz. Quando essa diversidade atravessa o limiar, ocorre um aumento da viremia, diminuição das células CD4 e progressão da doença (Nowak *et al.*, 1991).

## 1.8 – ENVELOPE

### 1.8.1 – GWGR *versus* GPGR

No Brasil, desde o início da década de 90, isolados da variante GWGR foram encontrados nas amostras do subtipo B (Potts *et al.*, 1993; Morgado *et al.*, 1994; Casseb *et al.*, 1998; Bongertz *et al.*, 2000; Vicente *et al.*, 2000). Muitas amostras do subtipo B brasileiro apresentam a tétrede GWGR no topo da alça V3 (B'') (Potts, 1993), enquanto a tétrede GPGR é encontrada na maioria das linhagens européia e norte-americanas (Bongertz *et al.*, 2000).

Cerca de 50% das variantes do subtipo B circulantes no Brasil formam um agrupamento genético e antigênico distinto (Hendry *et al.*, 1996). Essa variante possui como característica a codificação da seqüência de aminoácidos GWGR (ao invés do comum GPGR) na coroa da alça V3 (Diaz *et al.*, 1997). O aminoácido prolina, codificado pelo códon CCA, presente no subtipo B (GPGR) é substituído pelo aminoácido triptofano codificado pelo códon TGG. Possui, portanto, sítio de glicosilação relacionado ao N nas posições 301, 302 e 303 o que pode alterar o reconhecimento da gp120 pelo sistema imune permitindo o escape do vírus devido ao não reconhecimento de epítomos que estarão ocultos pelo carboidrato.

A seqüência GPGR (aminoácidos 310, 311, 312 e 313) localizada no topo (ou coroa) da alça é um domínio conservado, enquanto as seqüências próximas às cisteínas, na base da alça, são relativamente variadas (De Jong *et al.*, 1992). Na extremidade dessa alça existe um tetrapeptídeo, que é utilizado para a subtipagem do HIV-1, através de análises das seqüências dos genes do envelope viral dos diferentes subtipos. Hoje em dia, pela elevada incidência de formas recombinantes circulantes, a subtipagem do HIV-1 é também realizada por meio de análises do genoma completo (Sá Filho *et al.*, 2005). As seqüências desses genes podem ser utilizadas para a construção de árvores filogenéticas em que se agrupam isolados do mesmo grupo (Arens, 1999).

Embora seja um domínio conservado, variantes da seqüência GPGR (característico do subtipo B) já foram descritas. Segundo estudo realizado por Morgado e colaboradores em 1998,



37% das seqüências correspondia a típica variante brasileira desse subtipo contendo a téttrade GWGR no topo da alça V3 (Morgado *et al.*, 1998).

A diferença de aproximadamente 72% em relação aos aminoácidos entre as seqüências GWGR e GPGR nas regiões V1-V2 sugerem que os isolados virais podem diferir substancialmente em suas propriedades biológicas (Da Costa *et al.*, 1995). Logo, pressões adaptativas talvez possam ter ocorrido durante a disseminação do HIV-1 no Brasil e então talvez essa variante tenha um antecessor filogenético comum.

Estudo prospectivo com uma coorte realizado no Brasil sugeriu uma taxa de queda mais rápida nas células T com uma progressão mais rápida à aids e morte entre os indivíduos infectados com a téttrade GPGR quando comparado com a téttrade GWGR (Santoro-Lopes *et al.*, 2000).

Estudos também demonstraram que mudanças em duas regiões flanqueando a seqüência do tetrapeptídeo no topo da região V3 do *env* são suficientes para mudar o tropismo viral frente a diferentes células T (vírus CXR4) (Fouchier *et al.*, 1992). Levando-se em conta essa variabilidade, são usados recombinantes moleculares para mostrar que determinantes do tropismo do vírus alojam-se dentro da região V3 do *env* e, conseqüentemente, a característica heterogênea de V3 pode representar diferentes propriedades fenotípicas, classificadas em dois estados distintos: um estado no início da infecção e outro no decorrer da doença (Milich *et al.*, 1993).

Muitos estudos mostraram ser o domínio V3 o principal determinante na formação de sincício (De Jong *et al.*, 1992, Milich *et al.*, 1993), embora estudos de clones moleculares de provírus não indutores de sincício (NIS) e indutores de sincício (IS) tenham demonstrado que as regiões V1, V2 e V4 também podem estar envolvidas nesse processo (Fouchier *et al.*, 1992; Groenink *et al.*, 1993). Análise de variantes naturais do HIV-1 e mutações pontuais introduzidas na alça V3 em clones moleculares do vírus mostraram que aminoácidos básicos (K ou R) na posição 306 e/ou 320 conferem ao vírus o fenótipo IS enquanto aminoácidos ácidos (D ou E) nessas mesmas posições conferem o fenótipo NIS e essa correlação entre o genótipo do V3 e o fenótipo viral é independente das regiões V1, V2 e V4 (Chesebro *et al.*, 1991; De Jong *et al.*, 1992, Milich *et al.*, 1993). Também foi observado que o fenótipo X4 está associado a uma carga média maior ou igual a 5 nas somatórias das cargas entre as duas cisteínas, enquanto carga inferior a 5 corresponde ao fenótipo R5 (Fouchier *et al.*, 1992).

Primeiramente o HIV-1 era classificado, de acordo com seu comportamento em culturas de células, em dois distintos grupos baseados na cinética de crescimento, citopatologia e tropismo celular. Desse modo, um determinado fenótipo poderia ser IS rápida/alta (para cinética e replicação) e trópica por células T ou NIS lenta/baixa e trópica por macrófagos, sendo esses os “comportamentos” mais freqüentes em variantes virais. Atualmente, com a descoberta da

utilização dos receptores de quimiocinas como co-receptores de fusão e entrada do vírus na célula (D'Souza & Harden, 1996), passou-se a descrever o fenótipo viral de acordo com o co-receptor. Sendo assim, variantes que se utilizam do co-receptor de quimiocina  $\alpha$  CXCR4 quando infectam linfócitos T CD4, são classificadas como X4 e variantes que se utilizam do co-receptor de quimiocina  $\beta$  CCR5 quando infectam macrófagos são classificadas R5. Variantes X4 são conhecidas por serem T trópicas, rápida/alta, indutoras de sincício (IS) ou MT-2 trópicas, enquanto variantes R5 são M trópicas, lenta/baixa, não indutoras de sincício (NIS) ou não MT-2 trópicas. Essas características biológicas *in vitro* são encontradas entre todas as variantes do HIV-1.

A razão da escolha da região *env* do HIV-1 para o estudo também consiste no fato de que essa é uma região imunodominante, tem relação com o tropismo celular, escape imune e citopatogenicidade (De Jong *et al.*, 1992). Logo, a região do envelope, além de sempre ser importante para a determinação de subtipos virais e estudos de vacinas, poderia ser utilizada também como marcador de progressão para a aids. Assim, a comparação das pressões impostas pelos hospedeiros sobre as amostras GWGR e sobre as amostras GPGR contribui para avaliar se existe diferença seletiva entre elas.

Atualmente, não se sabe se as amostras GWGR são uma linhagem distinta ou apenas variantes de topo da alça V3 do gene *env*, pois essa é uma região muito variável e que possui sinal filogenético grande. Dessa forma, dados gerados a partir do seqüenciamento do gene *env* do HIV-1 funcionam como um excelente banco de informações. A utilização de metodologias menos sensíveis pode, eventualmente, apresentar artefatos que superestimem os resultados encontrados. Portanto, isso demanda a necessidade de um estudo que aborde essa questão, pois possibilitaria melhores progressos no acompanhamento do indivíduo infectado com HIV-1.

Alguns estudos sugerem que a presença de resposta citotóxica direcionada aos diferentes epítomos do gene *gag* esteja associada a uma progressão lenta para aids (Riviere *et al.*, 1995). Por outro lado, estudos realizados *in vitro*, evidenciaram que a resposta de CTL está direcionada principalmente contra poucos epítomos dos genes *env* e *pol*, sugerindo que a progressão tende a ser rápida (Hay *et al.*, 1999). Esses dados incentivaram-nos a expandir este estudo, além do gene *env*, para os genes *pol* e *gag*, pois explorar grandes trechos de seqüências virais de outras regiões para verificar se a seqüência GWGR é uma assinatura da alça V3 ou apenas uma variante, possibilitariam a obtenção de dados mais precisos e fidedignos que permitam um melhor entendimento da epidemia HIV/aids no Brasil.

## 2.0 – OBJETIVOS

### 2.1 – OBJETIVO PRINCIPAL:

- Verificar se as amostras GWGR constituem uma linhagem genética própria ou são apenas variantes de topo da alça V3 do gene *env* do HIV-1.

### 2.2 – OBJETIVOS SECUNDÁRIOS:

- Localizar amostras GWGR e construir um banco de amostras;
- Estudar a região *env* do HIV-1 com busca de assinaturas da alça V3 para GWGR e para GPGR e análise filogenética das seqüências;
- Explorar grandes trechos de seqüências virais de outras regiões do genoma do HIV-1 para verificar se amostras GWGR possuem apenas uma assinatura da alça V3 ou são uma variante do subtipo B do HIV-1;
- Avaliar a diversidade genética das amostras GWGR e comparar com seqüências GPGR;
- Comparar as pressões impostas pelos hospedeiros sobre as amostras GWGR e GPGR para avaliar a existência de pressão seletiva diferenciada.

### **3.0 – CASUÍSTICA**

#### **3.1 – ASPECTOS ÉTICOS**

O presente estudo foi submetido à apresentação do Comitê de Ética em Pesquisa do Hospital São Paulo/UNIFESP e aprovado sob o nº 1207/07.

#### **3.2 – CRITÉRIOS DE INCLUSÃO NO ESTUDO**

Foram considerados os seguintes critérios para a inclusão dos pacientes nos diferentes grupos:

##### **Grupo 1**

1. Ser infectada pelo subtipo B do HIV-1;
2. Possuir assinatura GWGR.

##### **Grupo 2**

1. Ser infectada pelo subtipo B do HIV-1;
2. Possuir assinatura GPGR.

#### **3.3 – SELEÇÃO DAS AMOSTRAS**

Inicialmente foi realizada uma triagem de amostras partindo-se de dados do seqüenciamento da alça V3 do gene *env* do HIV-1 gerados previamente no Laboratório de Retrovirologia da UNIFESP para selecionar amostras com a seqüência GWGR. A partir da triagem realizada foi construído um banco de amostras, a fim de obter-se a relação das amostras GWGR disponíveis para o estudo. Foram utilizadas todas as amostras disponíveis, não sendo realizado cálculo estatístico. Assim sendo, o número de amostras foi ajustado segundo o que se conseguiu obter.

Foram então separadas 80 amostras GWGR, sendo inicialmente utilizadas para o presente estudo 60 delas. O número maior de amostras GWGR separadas em relação ao número total de amostras a serem estudadas (60) ocorreu para que aquelas que eventualmente apresentassem

problemas na amplificação pudessem ser substituídas, a fim de garantir qualidade aos dados obtidos.

A fim de enriquecer o trabalho foram acrescentadas 40 amostras GPGR (sendo inicialmente utilizadas para o presente estudo 32 delas) partindo-se de dados do seqüenciamento da alça V3 do gene *env* do HIV-1 gerados previamente no Laboratório de Retrovirologia da UNIFESP para selecionar amostras com a seqüência GPGR. A partir da triagem realizada foi construído um banco de amostras, com intuito de obter-se a relação das amostras disponíveis para o estudo. O número maior de amostras GPGR separadas em relação ao número total de amostras a serem estudadas (32) ocorreu para que aquelas que eventualmente apresentassem problemas na amplificação pudessem ser substituídas, buscando garantir qualidade aos dados obtidos.

Portanto, foram inicialmente utilizadas para o presente estudo 60 amostras GWGR e 32 amostras GPGR, a partir do sangue total, que estavam disponíveis no Laboratório de Retrovirologia da Escola Paulista de Medicina – UNIFESP, São Paulo – SP previamente identificadas e colhidas em tubos contendo etilenodiaminotetracetato (EDTA). Essas amostras positivas para o HIV-1 (com sorologia confirmada através de ensaio imunoenzimático (ELISA) e *Western Blot*) são provenientes de amostras de indivíduos estocadas em freezers oitenta graus negativos no Laboratório de Retrovirologia – UNIFESP de indivíduos que já haviam iniciado o tratamento com terapia anti-retroviral.

## 4.0 – METODOLOGIA

### 4.1 – EXTRAÇÃO DO ÁCIDO NUCLEÍCO DO HIV-1

A extração do DNA de 80 amostras GWGR e 40 amostras GPGR do creme leucocitário (*buffy coat*) congeladas a  $-20^{\circ}\text{C}$  foi realizada utilizando-se o QIAamp DNA Blood Mini Kit (Quiagen Inc. Santa Clarita, CA, EUA). O número maior de amostras extraídas em relação ao número total de amostras (60 para GWGR e 32 para GPGR) para o estudo ocorreu para que as amostras que apresentassem problemas na amplificação pudessem ser substituídas, a fim de garantir qualidade aos dados obtidos. O *kit* realiza a extração do DNA por meio de quatro etapas:

- Reagente de lise celular (protease) para liberar o DNA das células mononucleares do sangue periférico (PBMC);
- Tampão (AL) e etanol absoluto para otimizar a ligação do DNA na membrana contida na coluna oferecida pelo *kit*;
- Tampões de lavagem (AW1 e AW2) para retirar os resíduos da reação;
- Tampão de eluição (AE) para retirar o DNA extraído da membrana.

Resumidamente, em tubo cônico de 1,5 mL foram transferidos 200  $\mu\text{L}$  do “creme leucocitário” e, adicionados 20  $\mu\text{L}$  de protease e 200  $\mu\text{L}$  solução tampão de lise (AL, com baixo pH e alta concentração de sais), incubados a  $56^{\circ}\text{C}$  por 10 minutos. Esse procedimento é necessário para que haja o rompimento das membranas celulares e nucleares, além da denaturação das proteínas presentes nas células, a fim de que seja liberado o DNA genômico. Sob condições de baixo pH adicionadas a etanol absoluto foi possível a adsorção do DNA à coluna de sílica quando centrifugados por 1 minuto à 16.000 rpm. Com o DNA “preso” à malha da sílica, foram realizadas lavagens sucessivas para a remoção de hemácias, proteínas e outras substâncias não-DNA. Para finalizar, o DNA foi eluído com 50  $\mu\text{L}$  de solução tampão com alto pH e baixa concentração de sais e armazenado em freezer  $-20^{\circ}\text{C}$ .

### 4.2 – REAÇÃO EM CADEIA PELA POLIMERASE (PCR) EM DUAS ETAPAS (NESTED) PARA A AMPLIFICAÇÃO DO DNA PROVIRAL

Após a extração do ácido nucleíco do HIV-1, foi utilizada a técnica de *nested*-PCR para amplificar os genes *env*, *pol* e *gag* do HIV-1. Essa técnica é uma modificação da reação em cadeia

pela polimerase (PCR). Consiste em duas etapas da PCR, envolvendo dois pares de iniciadores (*primers*). Na primeira etapa da reação um segmento de DNA é amplificado com um par de *primers*. Uma alíquota do material amplificado é então submetida à nova amplificação usando-se novo par de *primers*, localizados internamente em relação à posição do par de *primers* inicialmente utilizados. A segunda amplificação com *primers* diferentes, garante a especificidade da reação. Essa técnica apresenta sensibilidade ainda superior à da PCR e é recomendada quando a concentração de material nucléico é extremamente reduzida.

#### 4.2.1 – GENE *env*

Após a extração de DNA foi realizada a *nested-PCR* das amostras pertencentes aos grupos 1 e 2 utilizando um novo procedimento operacional padrão (POP) do Laboratório de Retrovirologia da UNIFESP para amplificar gene *env*.

A primeira etapa da *nested-PCR* amplifica um fragmento de 1411 pares de bases (bp), utilizando uma solução *master mix*, onde as concentrações dos reagentes utilizados para 1 mL da reação de PCR foram: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5mM, dNTP 0,2mM, iniciadores 0,2 pmol e Taq DNA polimerase 0,5 U.

<i>Primers</i>	Posição (HXB2)	Seqüência
ED 5	6557 – 6582	5' atg gga tca aag cct aaa gcc atg tg 3'
ED 12	7936 – 7968	5' agt get tcc tgc tgc tcc caa gaa ccc aag 3'

À essa solução adiciona-se 5µL do DNA extraído completando um volume final de 50µL. A reação foi incubada em um termociclador (ABI 9700 - Applied Biosystems, CA, EUA) e as condições de ciclagem foram:

94°C -----	5'	} 35 ciclos
94°C -----	30''	
55°C -----	45''	
72°C -----	1'	
72°C -----	5'	

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídio. A primeira etapa da reação é então concluída.

Para a segunda etapa, foram transferidos 5 $\mu$ L do produto amplificado na primeira reação para microtubos com uma solução *master mix* idêntica a acima descrita, mas adicionando nessa etapa 0,2 pmol de cada um dos iniciadores (*primers*) internos:

<i>Primers</i>	Posição (HXB2)	Seqüência
ES 7	6992 – 7021	5' tgt aaa acg acg gcc agt ctg tta aat ggc agt cta gc 3'
ES 8	7683 – 7648	5' cag gaa aca gct atg acc cac ttc tcc aat tgt ccc tca 3'

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídio. As condições físicas para essa reação foram as mesmas descritas na primeira etapa. Ao término dessa etapa obteve-se um produto de amplificação de 655 bp.

#### 4.2.2 – GENE *pol*

Após a extração de DNA foi realizada a *nested*-PCR das amostras pertencentes aos grupos 1 e 2, utilizando um novo procedimento operacional padrão (POP) do Laboratório de Retrovirologia da UNIFESP, porém já padronizado para amplificar a região da protease e da transcriptase reversa do gene *pol*.

A primeira etapa da *nested*-PCR amplifica um fragmento de 1913 pb, utilizando uma solução *master mix*, onde as concentrações dos reagentes utilizados para 1 mL da reação de PCR foram: Tampão 1X Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5mM, dNTP 0,2mM, iniciadores 0,2 pmol e Taq DNA polimerase 0,5 U.

<i>Primers</i>	Posição (HXB2)	Seqüência
POL 1	1842 – 1865	5' ggg agt ggg ggg acc cgg cca taa 3'
IMBO 1R	3739 – 3755	5' cca ctc agg aat cca ggt gg 3'



À essa solução adiciona-se 5 $\mu$ L do DNA extraído completando um volume final de 50 $\mu$ L. A reação foi incubada em um termociclador (ABI 9700 - Applied Biosystems, CA, EUA) e as condições de ciclagem foram:

94°C ----- 10'  
 94°C ----- 30''  
 55°C ----- 30'' } 35 ciclos  
 72°C ----- 2'  
 72°C ----- 10'

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídio. A primeira etapa da reação é então concluída.

Para a segunda etapa, foram transferidos 5 $\mu$ L do produto amplificado na primeira reação para microtubos com uma solução *master mix* idêntica a acima descrita, mas adicionando dessa vez 0,2 pmol de cada um dos iniciadores (*primers*) internos:

<i>Primers</i>	Posição (HXB2)	Seqüência
K1	2147 – 2166	5' cag agc caa cag ccc cac ca 3'
K2	5776 – 5745	5' ttt ccc cac taa ctt ctg tat gtc att gac a 3'

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídio. As condições físicas para essa reação foram as seguintes:

94°C ----- 10'  
 94°C ----- 45''  
 55°C ----- 45'' } 35 ciclos  
 72°C ----- 2'  
 72°C ----- 10'

Ao término dessa etapa obteve-se um produto de amplificação de 1200 bp.

### 4.2.3 – GENE *gag*

Após a extração de DNA foi realizada a *nested*-PCR das amostras pertencentes aos grupos 1 e 2, utilizando um novo procedimento operacional padrão (POP) do Laboratório de Retrovirologia da UNIFESP, porém já padronizado para amplificar o gene *gag*.

A primeira etapa da *nested*-PCR amplifica um fragmento de 1926 pb, utilizando uma solução *master mix*, onde as concentrações dos reagentes utilizados para 1 mL da reação de PCR foram: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5mM, dNTP 0,2mM, iniciadores 0,2 pmol e Taq DNA polimerase 0,5 U.

<b>Primers</b>	<b>Posição (HXB2)</b>	<b>Seqüência</b>
GAG A+	672 – 692	5' cca gag gag atc tct cga cgc 3'
DP 11	2598 – 2572	5' cca ttc ctg gct tta att tta ctg gta 3'

À essa solução adiciona-se 5µL do DNA extraído completando um volume final de 50µL. A reação foi incubada em um termociclador (ABI 9700 - Applied Biosystems, CA, EUA) e as condições de ciclagem foram:

94°C -----	5'	} 35 ciclos
94°C -----	45''	
58°C -----	45''	
72°C -----	2,5'	
72°C -----	5'	

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídio. A primeira etapa da reação é então concluída.

Para a segunda etapa, foram transferidos 5µL do produto amplificado na primeira reação para microtubos com uma solução *master mix* idêntica a acima descrita, mas adicionando dessa vez 0,2 pmol de cada um dos iniciadores (*primers*) internos:

<b>Primers</b>	<b>Posição (HXB2)</b>	<b>Seqüência</b>
SET III FW	896 – 913	5' ggg caa gca ggg agc tag 3'
CLIV N2	2268 – 2249	5' aaa gag tga tct gag gga ag 3'

Como controle negativo da reação de PCR foi utilizada água deionizada e como controle positivo foi usado o genoma do HXB2 clonado em plasmídeo. As condições físicas para essa reação foram as seguintes:

94°C -----	5'	} 35 ciclos
94°C -----	30''	
50°C -----	30''	
72°C -----	2'	
72°C -----	5'	

Ao término dessa etapa obteve-se um produto de amplificação de 1372 bp.

### 4.3 – DETECÇÃO DOS PRODUTOS AMPLIFICADOS

A fim de verificar a presença do produto amplificado com o tamanho esperado para as regiões mencionadas dos genes *env*, *pol* e *gag* dos grupos 1 e 2 foi realizado o fracionamento eletroforético utilizando 5 µL do produto de *nested*-PCR com 2 µL de tampão de carregamento (40% de sacarose; 0,25% de azul de bromofenol), que constitui na separação em gel de agarose 1,5% em tri-borato/ EDTA (TBE) 0,5X (0,89 M Tris; 0,89 ácido bórico e 0,02 M EDTA pH 8,0), contendo brometo de etídeo (0,5 µg/mL). A corrida foi efetuada a 100 Volts (V), no tempo médio de 60 minutos, em tampão TBE 0,5X. Para a comparação do peso molecular dos fragmentos amplificados foi utilizado um marcador de 100 bp (pares de bases) (GIBCO/BRL, Gaithersburg, MD, EUA). O gel foi visualizado sob iluminação ultravioleta (U.V.) (320 nm) e foi fotografado utilizando fotodocumentador GelDoc-It™ Imaging System (UVP, Inc. Cambridge, CA, EUA).

### 4.4 – PURIFICAÇÃO DOS PRODUTOS AMPLIFICADOS

Os produtos da PCR das amostras amplificadas dos genes *env*, *pol* e *gag* dos grupos 1 e 2 foram purificados utilizando o *kit* ChargeSwitch PCR Clean-up Kit (Invitrogen, Carlsbad, CA, EUA), conforme instruções do fabricante.

- Reagentes:
  - *Beads* magnéticos (25mg/mL em 10mM MÊS, pH 5.0, 10mM NaCl, 01% tween 20);
  - Tampão de purificação (N5);
  - Tampão de lavagem (W12);

- Tampão de eluição (E5; 10mM Tris-HCl, pH 8.5).

Esse *kit* de purificação de produto de PCR remove sais, iniciadores, dNTPs e outros reagentes não ácido-nucleicos. A tecnologia dos *beads* magnéticos é carga dependente auxiliada pelo tampão, que facilita a purificação do ácido nucleico. Em condições de baixo pH, os *beads* magnéticos têm uma carga positiva e se ligam ao ácido nucleico. Proteínas e outros contaminantes não se ligam e são lavados usando-se o tampão de lavagem. Para eluir o ácido nucleico, a carga na superfície é neutralizada pelo aumento de pH para 8,5 usando um tampão de eluição com baixa concentração de sal.

Resumidamente, em tubo cônico de 1,5 mL foram transferidos 50  $\mu$ L do produto de PCR e adicionados 50 $\mu$ L do tampão de purificação (N5) e 10 $\mu$ L dos *beads* magnéticos (previamente agitado), sendo homogeneizados com cuidado para não formar bolhas. Os tubos foram então incubados em temperatura ambiente por 1 minuto para ocorrer a ligação dos *beads* ao produto da PCR. Seguidamente, os tubos foram colocados na *Magna Rack* em frente ao ímã por 1 minuto ou até que se formasse o precipitado e, sem remover o tubo da frente do ímã, pipetado cuidadosamente o sobrenadante para não causar distúrbio ao precipitado formado para descartá-lo. Posteriormente, os tubos foram removidos da frente do ímã e foram adicionados 150 $\mu$ L de tampão de lavagem (W12), homogeneizando lentamente para dissolver o precipitado com o tampão cuidando para não formar bolhas. Os tubos foram colocados na *Magna Rack* em frente ao ímã por 1 minuto ou até que se formasse o precipitado e, sem remover o tubo da frente do ímã, pipetado cuidadosamente o sobrenadante, para não causar distúrbio ao precipitado formado para descartá-lo. Esse procedimento de lavagem foi novamente repetido. Por fim, foram adicionados 30 $\mu$ L do tampão de eluição (E5) e homogeneizados até dissolver o precipitado cuidando para não formar bolhas. Os tubos foram incubados em temperatura ambiente por 1 minuto, colocados na frente do ímã por 1 minuto ou até que se formasse o precipitado e, sem remover o tubo da frente do ímã, cuidadosamente transferido o sobrenadante, contendo o produto purificado da PCR para um tubo estéril (e guardado a menos – 20°C), sem causar distúrbio ao precipitado.

Após obter o produto de purificação, foi corrido um gel de agarose com 1  $\mu$ L de tampão de carregamento (40% de sacarose; 0,25% de azul de bromofenol) e 4  $\mu$ L do produto de purificação. Para a comparação do peso molecular dos fragmentos amplificados foi utilizado um marcador *Low DNA Mass Ladder* (Invitrogen, Carlsbad, CA, EUA). Como padrão foram aplicados 4 $\mu$ L desse marcador de massa molecular para comparar a quantidade de DNA da amostra (produto de amplificação) com o brilho do marcador de massa molecular (quanto maior a intensidade do brilho, maior a quantidade de DNA presente naquela amostra). A separação eletroforética foi

realizada em gel de agarose 1,5% (peso/volume) em Tris-borato/ EDTA (TBE) 0,5X (0,89M Tris; 0,89 M ácido Bórico e 0,02 M EDTA pH 8,0), contendo brometo de etídeo (0,5 µg/mL). A corrida foi efetuada a 100 Volts (V), no tempo médio de 60 minutos, em tampão TBE 0,5X. O gel foi visualizado sob iluminação ultravioleta (U.V.) (320 nm) e foi fotografado utilizando fotodocumentador GelDoc-It™ Imaging System (UVP, Inc. Cambridge, CA, EUA).

#### 4.5 – REAÇÃO DE SEQÜENCIAMENTO

Os produtos purificados da *nested*-PCR dos genes *env*, *pol* e *gag* dos grupos 1 e 2 foram submetidos à reação de seqüenciamento. Essa reação é similar a uma PCR, porém, de forma diferente da PCR, apenas um iniciador é utilizado para síntese das fitas, no seqüenciamento de DNA.

A quantificação do produto da PCR purificado em nanogramas (ng) por meio da sua comparação a um padrão de peso molecular *Low DNA Mass Ladder* (200mL, 4mL, Invitrogen, Califórnia, USA), observado em gel de agarose, foi previamente realizada, para determinar a quantidade de amostra utilizada na reação de seqüenciamento. A solução para cada amostra teve volume final de 6µL com água deionizada q.s.p, e o volume de cada amostra para essa solução foi definido por meio da quantificação realizada, como descrito acima, onde cada amostra deve conter aproximadamente 40ng.

Após a quantificação e diluição das amostras para 40ng, foi realizada a reação utilizando-se o kit *Amersham Biosciences DYEnamic ET Terminator Cycle Sequencing Kit* (Piscataway, EUA). As concentrações dos reagentes para seqüenciamento foram: *Terminator Ready Reaction Mix* (*Terminator Ready Reaction Mix* contém: dNTPs, ddNTPs, AmpliTaq DNA polimerase, MgCl<sub>2</sub>, tampão Tris-HCl) 2 µL, tampão 4 µL, iniciador 0,2 pMol, 3 µL (de cada iniciador), (para um volume final de 9 µL) e amostra. Cabe ressaltar que a solução para cada amostra teve um volume final de 6µL com água deionizada q.s.p, e o volume de cada amostra para essa solução foi definido por meio da quantificação realizada através da sua comparação a um padrão *Low DNA Mass Ladder* (200mL, 4mL/appl - Invitrogen, Califórnia, USA), observado em gel de agarose pós-purificação, onde cada amostra corresponde a aproximadamente 40ng, conforme já dito. A solução final totaliza de 15 µL por reação.

O seqüenciamento é determinado bi-direcionalmente, usando-se um conjunto de *primers*, sendo um na direção 5' → 3' (*sense*) e um na direção 3' → 5' (*antisense*). Os iniciadores utilizados para o seqüenciamento dos genes *env*, *pol* e *gag* foram:

<b>Genes</b>		
<b>Primers</b>	<b>Seqüência</b>	<b>Posição (HXB2)</b>
<b>Gene <i>env</i></b>		
LB <i>inner</i> 5'	5' gca gtc tag cag aag aag a 3'	7015 – 7030
LB <i>inner</i> 3'	5' tat gag gga caa ttg gag aag 3'	7666 – 7649
<b>Gene <i>pol</i></b>		
DP 10	5' caa ctc cct ctc aga agc agg agc cg 3'	2201 – 2223
F2	5' gta tgt cat tga cag tcc agc 3'	3308 – 3324
K1	5' cag agc caa cag ccc cac ca 3'	2147 – 2166
K2	5' ttt ccc cac taa ctt ctg tat gtc att gac 3'	5576 – 5745
<b>Gene <i>gag</i></b>		
SET III FW	5' ggg caa gca ggg agc tag 3'	896 – 913
SET III RV	5' ggg tgg ctc ctt ctg 3'	1330 – 1316
CLIV N1 FW	5' tgg tcc aaa atg cga acc	1754 – 1771
CLIV N2 RV	5' aaa gag tga tct gag gga ag 3'	2268 – 2249
A2 R	5' att gct tca gcc aaa act ctt gc 3'	1889 – 1867

A reação de seqüenciamento foi incubada em um termociclador (ABI 9700 - Applied Biosystems, CA, EUA) sob as condições:

94°C ----- 20''	} 25 ciclos
50°C ----- 15''	
60°C ----- 1'	

Durante essa reação o iniciador se liga à seqüência complementar e os nucleotídeos são incorporados de acordo com a fita molde. Ocorre tanto a incorporação dos nucleotídeos livres não marcados, quanto a dos nucleotídeos marcados com fluorescência. Esses nucleotídeos são chamados de nucleotídeos de terminação (marcados), pois, cada vez que um nucleotídeo de terminação é incorporado, a leitura da fita molde é interrompida, gerando um fragmento de tamanho respectivo ao local onde o nucleotídeo de terminação foi incorporado. No final da reação

são gerados fragmentos de tamanhos diferentes e também há sobra de *primers*, ddNTPs, alguns sais, entre outros resquícios.

#### 4.6 – PRECIPITAÇÃO DA REAÇÃO DE SEQÜENCIAMENTO

No final da reação de seqüenciamento, como já dito, são gerados fragmentos de tamanhos diferentes e também há sobra de iniciadores, ddNTPs, alguns sais, entre outros resquícios. Os produtos gerados nessa reação foram precipitados para eliminar os resíduos.

Resumidamente, foram adicionados ao produto do seqüenciamento 90µL de etanol 70%, e misturados em agitador automático. Após incubação de 15 minutos à temperatura ambiente, protegido da luz, foi realizada centrifugação a 4000 rpm por 45 minutos à 20° C. O resultado da centrifugação é a formação de um botão (*pellet*) que contém o produto do seqüenciamento precipitado. O sobrenadante foi então desprezado e, 150 µL de etanol 70% foram adicionados novamente. Uma segunda etapa de agitação e centrifugação a 4000 rpm por 15 minutos à 20° C foi submetida às amostras para nova lavagem e formação de botão (*pellet*) mais purificado. Após a retirada do sobrenadante, o produto precipitado foi colocado no termociclador até a temperatura de 94° C por 3 minutos, aberto, para secagem completa do botão (*pellet*).

As amostras precipitadas foram ressuspensas em 15µL de formamida *HiDi*, centrifugadas a 400 rpm por 1 minuto à 20° C. A seguir, foram denaturadas à 95°C por 3 minutos e seqüenciadas no seqüenciador automático ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, CA, EUA) para separação eletroforética.

#### 4.7 – ANÁLISE DO SEQÜENCIAMENTO

Os sinais fluorescentes gerados no seqüenciamento automático são transferidos para um computador que analisa a posição e a força do sinal, produzindo um eletroferograma. Por meio da análise do eletroferograma gerado a partir da leitura dos fragmentos de DNA no seqüenciamento automático foram realizadas as edições utilizando-se o programa Sequencer 4.8 (Gene Code Corporation, EUA).

A área sob o pico do eletroferograma representa a intensidade do sinal e a cor do pico é especificada de acordo com a base nucleotídica em questão. Sendo assim, o programa nomeia a

base A, C, T ou G para cada posição ou, então, N. Nos casos em que a determinação das bases não é clara (ex. múltiplos picos sobrepostos), a nomeação das bases foi realizada de forma arbitrária, dentre os quais os picos com maior intensidade são considerados. Assim, usam-se os códigos da IUPAC (*International Union of Pure and Applied Chemistry* - União Internacional de Química Pura e Aplicada), citados na tabela abaixo (**Tabela 8**). Se em uma mesma posição do eletroferograma, por exemplo, existem dois picos com cores designando as bases A e G, então, usa-se a letra R, na sequência de DNA, para designar a ambigüidade A/G.

**Tabela 8.** Código IUPAC

Código	Tradução
A	Adenina
C	Citosina
G	Guanina
T	Timina
B	C, G ou T
D	A, G ou T
H	A, C ou T
R	A ou G
Y	C ou T
K	G ou T
M	A ou C
S	G ou C
W	A ou T
N	Qualquer base
V	A, C ou G

As edições e análise dos fragmentos seqüenciados, como já dito, foram feitas através do programa Seqüencher 4.8 (Genes Code Corporation, EUA), que fornece uma seqüência consenso entre as seqüências de nucleotídeos alinhadas, geradas pelos iniciadores *sense e antisense*.

Posteriormente, as seqüências de nucleotídeos obtidas foram traduzidas para aminoácidos e então, alinhadas por meio do programa Clustal X (THOMPSON *et al.*, 1997) e ajuste manual usando o programa Se-AL (Rambaut *et al.*, 2000).



## 4.8 – ANÁLISE DOS SUBTIPOS E DETECÇÃO DE RECOMBINAÇÃO

Após o alinhamento, as seqüências selecionadas foram subtipadas por meio de duas metodologias, para que nenhuma amostra pertencente a outros subtipos não B fosse analisada.

Inicialmente, foi realizada uma triagem utilizando o programa “*HIV blast search*” ([http://HIV-web.lanl.gov/content/HIV-db/BASIC\\_BLAST/basic\\_blast.html](http://HIV-web.lanl.gov/content/HIV-db/BASIC_BLAST/basic_blast.html)), devido à rapidez da obtenção de uma identificação de subtipos.

Outra metodologia empregada foi o SIMPLOT (Salminen *et al.*, 1995) disponível no programa SIMPLOT versão 3.5.1. Essa metodologia consiste na avaliação filogenética por algoritmo de distância de subalinhamentos seqüenciais e sobrepostos ao longo das seqüências analisadas. Esse programa ainda permite determinar os pontos de recombinação (ponto de quebra) nas seqüências analisadas, pois obtém-se a informação do ponto em que há alternância de filogenia entre as seqüências. Nessa análise foram utilizadas janelas de 250 bp com incremento de 10bp para o gene *env*, janelas de 600 bp com incremento de 200 bp para o gene *gag* e janelas de 400 bp com incremento de 60 bp para o gene *pol*. A utilização e adequação desses parâmetros estão apresentadas nos resultados do presente trabalho.

As relações filogenéticas obtidas foram confirmadas utilizando-se o método de *Maximum Likelihood* (máxima verossimilhança) implementado pelo *software* PAUP\* (Swofford, 2002). Esse método utiliza matrizes de probabilidade de mudança entre nucleotídeos que explicitam modelos evolutivos para a inferência dos estados ancestrais dos caracteres, a partir dos estados observados nos táxons apicais.

Os pontos de recombinação foram analisados também pela inspeção manual das assinaturas dos nucleotídeos que compõem cada subtipo.

## 4.9 – ANÁLISE FILOGENÉTICA DAS SEQÜÊNCIAS OBTIDAS

### 4.9.1 – COMPILAÇÃO E ALINHAMENTO DAS SEQÜÊNCIAS

As seqüências foram alinhadas com referências obtidas de bancos de dados, utilizando o programa Clustal X (Thompson *et al.*, 1997). Esses alinhamentos foram utilizados para a

construção de árvores filogenéticas e para a detecção da pressão seletiva através do cálculo de dN/dS descrito a seguir.

#### 4.9.2 – INFERÊNCIA FILOGENÉTICA

As seqüências dos nucleotídeos obtidas pelo seqüenciamento dos genes *env*, *pol* e *gag* foram alinhadas com seqüências de nucleotídeos dos subtipos B, C e F principalmente, com o auxílio do programa Clustal X (THOMPSON *et al.*, 1997) e ajuste manual. A análise filogenética foi primeiramente realizada com a geração de árvores filogenéticas utilizando o pacote de programas PHYLIP 3.66 a partir de uma matriz de distância contendo 2 parâmetros através do método KIMURA. A seguir foi empregado o método de *Neighbor-Joining* (agrupamento por similaridade).

Em um momento mais avançado foi utilizado o método de inferência filogenética feito por máxima verossimilhança, o qual utiliza matrizes de probabilidade de mudança entre nucleotídeos que explicitam modelos evolutivos para a inferência dos estados ancestrais dos caracteres, a partir dos estados observados nos táxons apicais. As reconstruções ótimas maximizam a verossimilhança para todos os caracteres ao longo de uma topologia. No entanto, o número de substituições observadas em um alinhamento tende a subestimar o número real de eventos evolutivos para determinado conjunto de seqüências (Yang, 1994). Por isso, é necessário o uso de modelos evolutivos de substituição de nucleotídeos (ou aminoácidos) que permitem a correção do número de eventos evolutivos (substituições) num determinado conjunto de dados (seqüências). Os modelos utilizados para descrever o processo de evolução das substituições de nucleotídeos, ou transição de estados (*ex.* A→G) em seqüências de DNA (RNA), seguem as seguintes premissas: 1. as taxas de mutações para cada nucleotídeo não se alteram com o tempo; 2. disponibilidade celular de nucleotídeos é constante e 3. o estado atual de um nucleotídeo só depende do seu estado anterior, e não de estados intermediários. As probabilidades de transição (*P*) de nucleotídeos (T, C, A, G) de um estado (*i*) para outro estado (*j*) em um intervalo de tempo (*dt*), dadas as premissas anteriores, podem ser expressas de forma matricial (4x4).

$$P_i = \begin{pmatrix} P_{TT} & P_{CT} & P_{AT} & P_{GT} \\ P_{TC} & P_{CC} & P_{AC} & P_{GC} \\ P_{TA} & P_{CA} & P_{AA} & P_{GA} \\ P_{TG} & P_{CG} & P_{AG} & P_{GG} \end{pmatrix} \quad (1)$$

Esse modelo de representação matricial das probabilidades, expresso em termos de estados descontínuos, é conhecido como cadeia de Markov. Cada elemento  $p_{ij}$  da matriz corresponde a uma probabilidade de mudança de estado  $i$  pelo  $j$  (ex.  $A \rightarrow C$ ), num intervalo de tempo. Vários modelos evolutivos propostos diferem na complexidade dos elementos da matriz, indo de um modelo cujos elementos, em sua totalidade, têm a mesma probabilidade, 25% de transição (modelo Jukes e Cantor), até uma matriz mais geral, em que cada elemento tem uma probabilidade distinta.

Neste trabalho, a matriz de probabilidade de substituição foi escolhida de acordo com o modelo evolutivo indicado pelo teste de razão de verossimilhança LRT (“*Likelihood Ratio Test*”). O teste LRT é usado em inferências estatísticas e deve ser aplicado somente quando os modelos são aninhados. Dois modelos são aninhados quando o modelo mais complexo (aquele com um número maior de parâmetros livres) derivou de um modelo mais simples (com menos parâmetros livres) e, no mais complexo, existe a adição de pelo menos um parâmetro. Assim, após a estimativa dos valores de verossimilhança do modelo mais simples ( $L_0$ ) e do mais complexo ( $L_1$ ), eles são tratados no teste LRT como hipótese nula ( $H_0$ ) e hipótese alternativa ( $H_1$ ), respectivamente. Isso é feito pela diferença de valores do logaritmo da verossimilhança entre dois modelos aninhados na seguinte equação:

$$\Delta LRT = (2[\ln L_1 - \ln L_0]) \quad (2)$$

Os valores de  $\Delta LRT$  seguem uma distribuição qui-quadrado assintótica ( $\chi^2$ ), pela qual os graus de liberdade são dados pela diferença de parâmetros livres entre os modelos. A hipótese nula é descartada quando o valor do teste for menor que o nível de significância escolhido (valor de alfa). As estimativas dos parâmetros e de verossimilhança para 56 modelos aninhados foram feitas com o programa PAUP. O teste LRT foi executado pelo programa MODELTEST 3.06b (Posada, Crandall, 1998). A heterogeneidade das taxas de substituição entre os sítios foi incorporada através da distribuição gama ( $\Gamma$ ). Nessa forma de distribuição, o parâmetro alfa ( $\alpha$ ) e a proporção de sítios que não variam foram estimados a partir dos dados. Da mesma forma, os valores das matrizes de transição e a frequência de bases também foram estimados de forma interativa durante o processo de reconstrução. A reconstrução filogenética foi feita em etapas:

1. Iniciou-se com a escolha do modelo de evolução de nucleotídeos e obtenção de uma árvore inicial pelo método de distâncias de agrupamento de vizinhos (“*neighbor joining*”);

2. Posteriormente, foram feitos rearranjos (perturbações) na árvore inferida pelo método dos quartetos, através do método de busca heurística com o algoritmo “*stepwise addition*” e o método “*nearest neighbor interchange (NNI)*”. Ainda, em certos casos específicos, foi usado o método “*tree bisection and reconnection*” (TBR) de rearranjo das árvores. As árvores com maior verossimilhança que a inicial foram escolhidas.

Esse método de reconstrução filogenética é bastante apropriado para o ajuste do comprimento dos ramos, visto que, em cada uma das etapas, é feito um refinamento dos parâmetros da matriz de transição (modelo escolhido pelo MODELTEST).

No presente trabalho foram utilizadas matrizes de probabilidade de mudança entre nucleotídeos que explicitam modelos evolutivos para a inferência dos estados ancestrais dos caracteres, a partir dos estados observados nos táxons apicais. As reconstruções ótimas maximizam a verossimilhança (*maximum likelihood*) para todos os caracteres ao longo de uma topologia. As árvores de verossimilhança foram feitas utilizando o software PAUP\* (Swofford *et al.*, 2002).

### 4.9.3 – MÉTODOS DE DETECÇÃO DE SELEÇÃO

Para a detecção de seleção natural foi utilizada a razão entre o número de substituições não-sinônimas por sítio não-sinônimo (dN) e o número de substituições sinônimas por sítio sinônimo (dS). O regime de seleção foi caracterizado pela razão entre dN / dS ( $\omega$ ), onde:

$\omega > 1$  (dN > dS) para seleção positiva.

$\omega < 1$  (dN < dS) para seleção negativa ou purificadora.

$\omega = 1$  (dN = dS) para um regime neutro de substituições.

#### 4.9.3.1 - DETERMINAÇÃO DE DN E DS PELO MÉTODO DE DISTÂNCIA PAR-A-PAR

Foram calculadas as distâncias genéticas normalizadas pelo número de sítios ao longo das seqüências. As distâncias são computadas par-a-par para um alinhamento composto por  $n$  seqüências de DNA de mesmo comprimento (método de Nei-Gojobori).

Especificamente, no método de Nei-Gojobori,  $dN$  e  $dS$  são estimadas como a razão do número total de substituições sinônimas ( $S$ ) e não-sinônimas ( $N$ ), divididos respectivamente pelo número de sítios potencialmente sinônimos ( $s$ ) e potencialmente não-sinônimos ( $n$ ). Esses valores são computados para cada códon sob a premissa de que as mudanças de nucleotídeos têm a mesma probabilidade (modelo de Jukes-Cantor). Assumindo que  $f_i$  designa a proporção de mudanças sinônimas possíveis na posição  $i$ -ésima de um determinado códon ( $i=1, 2, 3$ ), o número de sítios potencialmente sinônimos ( $s$ ) pode ser obtido por:

$$s = \sum_{i=1}^3 f_i \quad (3)$$

O número de sítios potencialmente não-sinônimos ( $n$ ) é dado por  $n=3-s$ . Para exemplificar, temos o caso da trinca da fenilalanina (TTT), na qual as substituições, na primeira e na segunda posição, são não-sinônimas e, somente, 1 em 3 substituições são sinônimas (TTC). Assim, temos que o número de sítios potencialmente sinônimos ( $s$ ) é igual a:  $0+0+1/3$ . Por outro lado, o número de sítios potencialmente não-sinônimos ( $n$ ) é igual a:  $3-1/3=8/3$ . Quando uma mudança resulta num códon de terminação, este não é computado. Para obter os valores do número total de substituições sinônimas ( $S$ ) na seqüência inteira, usa-se a expressão:

$$S = \sum_{j=1}^c s_j \quad (4)$$

Para os valores do número de substituições não-sinônimas ( $N$ ), usa-se a expressão  $N=3C-S$ , na qual  $C$  é o número total de códons e  $s_j$  é o valor do número de sítios potencialmente sinônimos para o  $j$ -ésimo códon. Nota-se que  $S+N=3C$  é igual ao número total de nucleotídeos comparados. O cálculo das diferenças de nucleotídeos sinônimos e não-sinônimos é feito par-a-par, códon por códon. Por exemplo, na comparação dos pares de códons TTT (F) e GTA (V) existem duas possibilidades neste caminho:



Designando os símbolos  $s_d$  e  $n_d$  para as diferenças sinônimas e não sinônimas por códon, respectivamente, e assumindo que ambos os caminhos (1) e (2) têm as mesmas probabilidades, tem-se os valores  $s_d = 1/2$  e  $n_d = 3/2$ . Nos casos em que as substituições resultarem em códons de

terminação, os passos não são computados. Com base nessas premissas, é possível estimar a proporção das diferenças sinônimas ( $p_s$ ) e não-sinônimas ( $p_n$ ) por:  $p_s = s_d / S$  e  $p_n = s_n / N$ , respectivamente. Nesse contexto, S e N são os números médios de sítios sinônimos e não-sinônimos para duas seqüências. Para estimar o número de substituições sinônimas (dS) e o número de não sinônimas (dN) por sítio, usa-se a seguinte expressão, sugerida pelo modelo de Jukes-Cantor:

$$dX = -(3/4) \ln[1 - (4/3)p_x] \quad (5)$$

Segundo esse modelo, substitui-se  $p$  por  $p_s$  ou  $p_n$ . Assim, o número de substituições sinônimas é calculado por:  $dS = -(3/4) \ln[1 - (4/3)p_s]$  e o número de substituições não sinônimas é calculado por  $dN = -(3/4) \ln[1 - (4/3)p_n]$ . Com isso, o valor de  $\omega$  é calculado a partir das médias dos valores de  $dN$  e  $dS$ . Esse método fornece uma estimativa aproximada de  $\omega$  visto que, o processo de substituição nos sítios sinônimos e não sinônimos, não segue o modelo de Jukes-Cantor. Além de ser sensível ao comprimento e ao número de seqüências analisadas, outros fatores podem influenciar nos resultados. O excesso de polimorfismos e as taxas de heterogeneidade entre os sítios, também podem influenciar nas estimativas.

Vários programas usam o método de Nei-Gojobori (1986). Em nossas análises foi utilizado o programa MEGA 4.0, que calcula a média geral entre os valores de dN e dS e seus erros associados para todos os pares de seqüências.

#### 4.9.3.2 – DETERMINAÇÃO DE DN E DS PELO MÉTODO DE VEROSSIMILHANÇA MÁXIMA

Yang (1994) e Muse & Gaut (1994) propuseram uma estimação da relação entre as taxas de substituições não-sinônimas e sinônimas,  $\omega$ , dentro de um contexto da verossimilhança máxima (*maximum likelihood*), utilizando um modelo de substituição de códon. Nielsen e Yang (1998) propuseram um método segundo o qual os sítios (códon) de uma proteína podem assumir valores de dN e dS distintos para cada sítio. Com isso, a detecção de seleção positiva pode ser feita individualmente para cada um dos sítios (códon) para uma determinada proteína. Esse método está implementado no programa CODEML, do pacote PAML v.3.154 (Yang, 1997). O CODEML utiliza vários modelos evolutivos que diferem entre si na distribuição de  $dN/dS$  ( $\omega$ ) entre os códon.

Dentre os vários modelos de substituição de códons implementados no programa CODEML, alguns deles são mais ajustáveis às sequências de DNA de vírus e serão discutidos abaixo (Nielsen & Yang, 1998; Yang *et al.*, 1997):

1. o M0 é o mais simples, pois assume um único valor de dN/dS para todos os sítios (códons);
2. o M1 (neutro) assume duas categorias de códons: os conservados ( $dN/dS = 0$ ) e os neutros ( $dN/dS = 1$ ), distribuídos em diferentes proporções ( $p_0$  e  $p_1$ );
3. o M2 (selecção), comparado ao modelo anterior, adiciona uma nova categoria de códons ( $p_2$ ), na qual pode assumir valores maiores que 1 para dN/dS ( $\omega_2$ ) e;
4. o M3 (discreto) permite n categorias de códons para um determinado número de classes (K). Os valores de dN/dS ( $\omega$ ), por categoria, são estimados a partir dos dados e seguem uma distribuição gama, subdividida em K categorias. Em nossos estudos, utilizamos  $K=8$ , de acordo com Yang *et al.* (2000). Ressalta-se que foram obtidos valores de  $\omega_0$  até  $\omega_7$  e suas respectivas proporções ( $p_0$  até  $p_7$ ) de códons pertencentes às diferentes classes. A proporção  $p_8$ , de sítios com valores de dN/dS mais altos, avalia os códons com selecção positiva mais intensa.

A abordagem de detecção de selecção positiva é inicialmente feita pela estimativa de parâmetros para todos os modelos acima citados; a seguir, o teste LRT é usado para verificar qual o modelo que melhor se ajusta aos dados. O teste é feito pela diferença dos valores de verossimilhança entre os modelos aninhados. Essa diferença segue uma distribuição qui-quadrado assintótica, a qual permite determinar a significância estatística entre os modelos, assumindo que os graus de liberdade são dados pela diferença entre os parâmetros livres de cada modelo.

A detecção de selecção positiva ocorre se o modelo escolhido assumir valores de dN/dS maiores que um ( $\omega > 1$ ) e, também, se os valores estimados forem significativos. Assim, se o teste LRT indicar que o modelo M2 é mais adequado que os modelos M0 e M1, então a hipótese de selecção positiva é aceita. Se acaso o teste indicar que o modelo M3 é mais adequado, quando comparados os modelos M1 vs. M3 e M2 vs. M3, é permitida a identificação dos códons sob selecção positiva. A melhor comparação entre M0 vs. M2 e M1 vs. M2, é mais conservadora na estimativa para detectar selecção positiva (Yang, 1997).

Além disso, foi utilizado o método de verossimilhança, o qual avalia se o número de substituições não-sinônimas por sítio não-sinônimo ( $dN$ ) e o número de substituições sinônimas por sítio sinônimo ( $dS$ ) diferem entre duas populações (Pond *et al.*, 2005). A significância das diferenças é avaliada por LRT, assumindo um grau de liberdade.

A taxa de substituição não-sinônima/sinônima ( $\omega=dN/dS$ ) fornece uma medida sensível da pressão seletiva, pois possibilita a adequação de modelos de substituição de códons (uso de diferentes matrizes de transição de aminoácidos) e o ajuste das taxas de heterogeneidade nas estimativas de  $\omega$  para cada códon. Também, possibilita o uso da informação genealógica dos dados para descrever o processo evolutivo de substituição de códons (Felsenstein, 1973; Yang, 1994).



## 5.0 – RESULTADOS

### 5.1 – EXTRAÇÃO

A extração do DNA de 80 amostras GWGR e 40 amostras GPGR do creme leucocitário (*buffy coat*) congeladas a  $-20^{\circ}\text{C}$  foi realizada utilizando-se o QIAamp DNA Blood Mini Kit (Quiagen Inc. Santa Clarita, CA, EUA). O número maior de amostras extraídas em relação ao número total de amostras (60 para GWGR e 32 para GPGR) para o estudo ocorreu para que as amostras que apresentassem problemas na amplificação pudessem ser substituídas, a fim de garantir qualidade aos dados obtidos.

**Tabela 9: Quantidade de amostras extraídas, amplificadas e seqüenciadas para GWGR**

<b>Genes</b>	<b>Extração de DNA</b>	<b>PCR</b>	<b>Seqüenciamento</b>
<i>Env</i>	80	65	60
<i>Pol</i>	80	65	60
<i>Gag</i>	80	65	60

Das 80 amostras extraídas para GWGR, 65 apresentaram amplificação positiva. As amostras amplificadas foram então seqüenciadas. Das 65 amostras seqüenciadas, 5 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 60 amostras para GWGR.

**Tabela 10: Quantidade de amostras extraídas, amplificadas e seqüenciadas para GPGR**

<b>Genes</b>	<b>Extração de DNA</b>	<b>PCR</b>	<b>Seqüenciamento</b>
<i>Env</i>	40	35	32
<i>Pol</i>	40	35	32
<i>Gag</i>	40	35	32

Das 40 amostras extraídas para GPGR, 35 apresentaram amplificação positiva. As amostras amplificadas foram então seqüenciadas. Das 35 amostras seqüenciadas, 3 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 32 amostras para GPGR.

## 5.2 – EXTRINGÊNCIA DA REAÇÃO

### 5.2.1 – GENE *env*

O produto amplificado a partir da *nested*-PCR com os iniciadores LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5mM, dNTP 0,2mM, iniciadores 0,2 pmoL e Taq DNA polimerase 0,5 U, sob as mesmas condições de ciclagem descritas anteriormente, e posteriormente visualizado pela separação eletroforética em gel de agarose, apesar de obter uma grande quantidade de produto amplificado do tamanho esperado, apresentou uma grande quantidade de bandas inespecíficas altamente evidentes (**Figura 6**), tanto nas amostras como no controle positivo, o que sugere um problema na reação que poderia intervir na etapa de seqüenciamento. Portanto, foi necessária a realização da padronização da *nested*-PCR para o gene *env*.

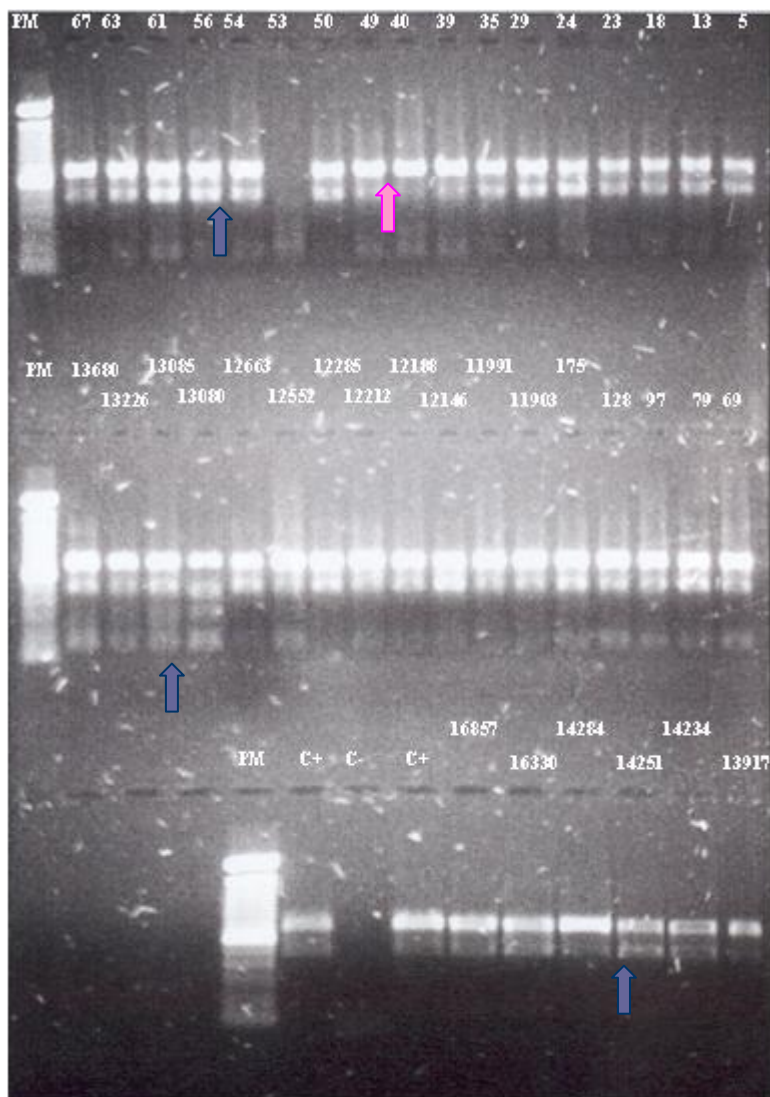
Para a padronização da *nested*-PCR foram utilizados testes para verificar qual a melhor concentração de reagentes para a reação, para tanto foram testadas as concentrações de magnésio, de dNTPs, de iniciadores e da enzima (**Figura 7**) nas seguintes condições:

1. Iniciadores LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), **MgCl<sub>2</sub> 2,5 mM**, **dNTP 0,2 mM**, iniciadores 0,2 pmoL e Taq DNA polimerase 0,5 U, sob as mesmas condições de ciclagem descritas anteriormente;

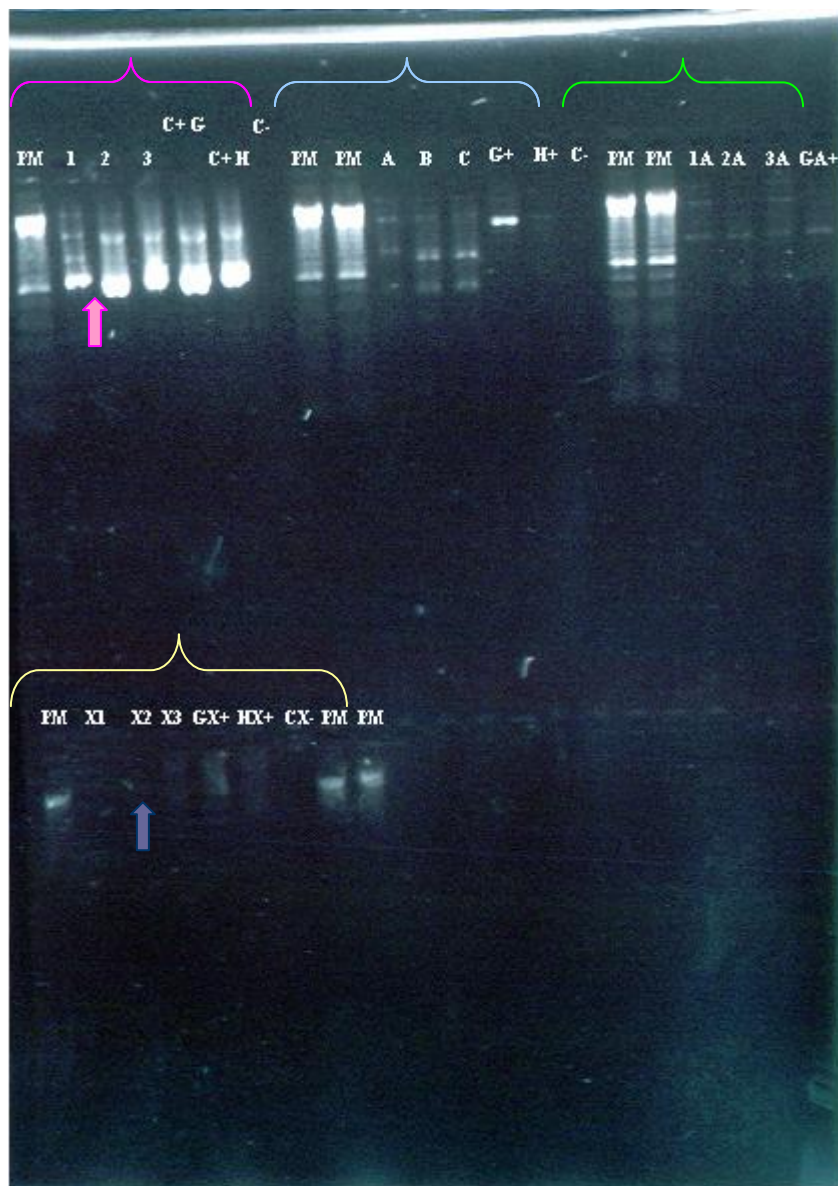
2. Iniciadores LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), **MgCl<sub>2</sub> 3,5 mM**, **dNTP 0,4 mM**, iniciadores 0,2 pmoL e Taq DNA polimerase 0,5 U, sob as mesmas condições de ciclagem descritas anteriormente;

3. Iniciadores LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5 mM, dNTP 0,2 mM, **iniciadores 0,4 pmoL** e **Taq DNA polimerase 0,25 U**, sob as mesmas condições de ciclagem descritas anteriormente;

4. Iniciadores LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), **MgCl<sub>2</sub> 3,5 mM**, **dNTP 0,4 mM**, **iniciadores 0,4 pmoL** e **Taq DNA polimerase 0,25 U**, sob as mesmas condições de ciclagem descritas anteriormente.



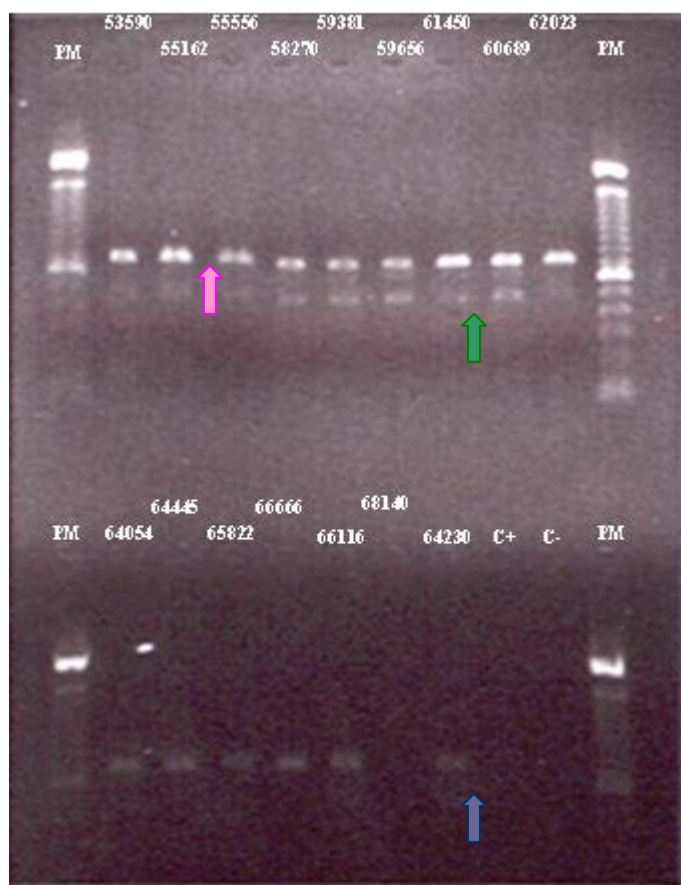
**Figura 6.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença de bandas inespecíficas. As setas azuis indicam inespecificidade e a seta rosa a banda alvo. PM significa marcador de peso molecular de 100bp. Nas três fileiras do gel encontram-se identificadas as amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo.



**Figura 7.** Eletroforese em gel de agarose 1,5% com controles para teste de concentração de magnésio, dNTP e iniciadores. A seta azul indica o pior resultado e a seta rosa a melhor banda no gel. PM significa marcador de peso molecular de 100bp. Nas duas fileiras do gel encontram-se identificadas os controles utilizados no teste. C+ é controle positivo e C- é controle negativo. A chave rosa indica a condição 1, a chave azul a condição 2, a chave verde a condição 3 e a chave amarela a condição 4.

Posteriormente, foram realizados testes que compreendem a variação da temperatura de anelamento dos iniciadores de 55°C e 58°C, e o resultado dessa reação pode observado na foto da eletroforese (**Figura 8**). Quando comparado ao anterior, esse resultado demonstrou uma pequena redução das bandas inespecíficas, contudo, as inespecificidades presentes ainda apresentavam

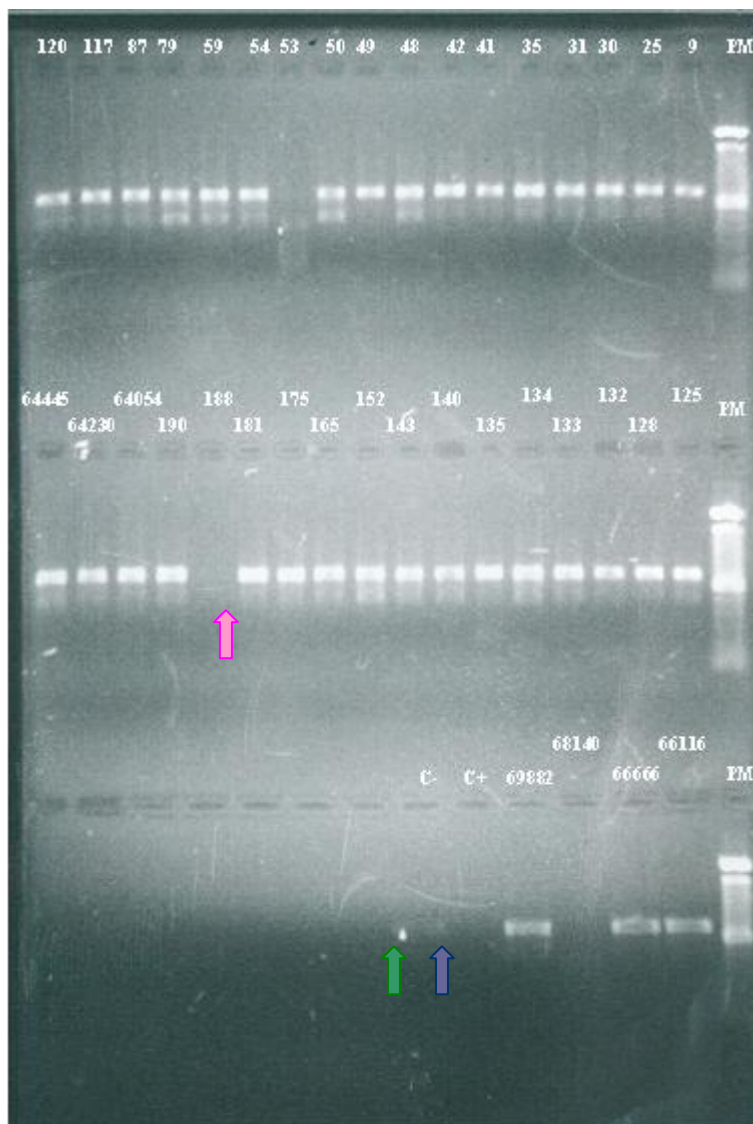
grande evidência; além disso, o controle positivo apresentou uma banda alvo muito fraca (permitindo sua visualização apenas quando o gel estava no transiluminador).



**Figura 8.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença de bandas inespecíficas em quantidades menores, mas ainda com grande intensidade (seta verde) e banda alvo fraca (seta rosa), além de controle positivo não observado na foto (seta azul). PM significa marcador de peso molecular de 100bp. Nas duas fileiras do gel encontram-se identificadas as amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo. Na primeira fileira a condição é de 55° C na temperatura de anelamento e na segunda fileira a condição é de 58° C na temperatura de anelamento.

A fim de aumentar a extrinsecência *nested*-PCR e obter um produto amplificado mais específico e com maior quantidade de DNA na banda alvo, foram realizados novos testes para ajustar a reação (porém, o controle positivo novamente apresentou uma banda alvo muito fraca

(permitindo sua visualização apenas quando o gel estava no transiluminador) e uma banda inespecífica no controle negativo), o que colocou ainda mais em dúvida a qualidade da reação, fazendo-se necessário outro teste (**Figura 9**). As condições foram: nova alíquota dos iniciadores LB *outer* (*sense* e *antisense* - 1ª etapa) e LB *inner* (*sense* e *antisense* - 2ª etapa) sob as condições de concentrações dos reagentes utilizados para 1 mL: Tampão 1X (200mM Tris-HCl pH 8,4 e 500mM KCl), MgCl<sub>2</sub> 2,5 mM, dNTP 0,2 mM, iniciadores 0,2 pmol e Taq DNA polimerase 0,5 U, e condições de ciclagem descritas anteriormente, além do aumento da quantidade de DNA para 10mL em cada tubo).



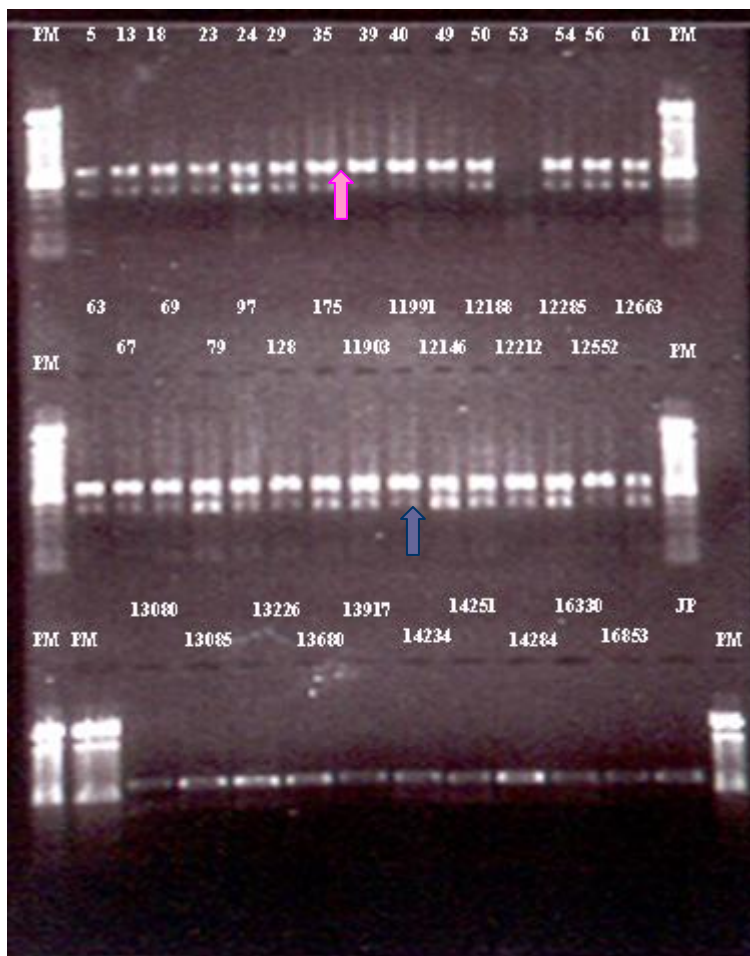
**Figura 9.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença, porém reduzida de bandas inespecíficas (seta rosa). A seta azul indica o controle positivo que não permite visualização pelo gel e a seta verde indica inespecificidade de banda no controle negativo. Teste com novas alíquotas dos iniciadores e maior quantidade de DNA na reação. PM significa marcador de peso molecular de 100bp. Nas três fileiras do gel encontram-se identificadas as amostras utilizadas no teste. C+ é controle positivo

Após o teste de amplificação realizado com as condições descritas na metodologia para o gene *env*, mantendo apenas os então iniciadores dos testes de amplificação para LB *outer* (*sense* e *antisense*) (1ª etapa) e LB *inner* (*sense* e *antisense*) (2ª etapa), o produto da PCR foi separado eletroforeticamente em gel de agarose (**Figura 10**) e, apesar da presença de bandas inespecíficas, seqüenciado para verificar se os iniciadores flanqueavam a região de interesse do estudo. Mesmo obtendo a confirmação da região de interesse, o resultado do seqüenciamento mostrou que os iniciadores da amplificação utilizados influenciavam no seqüenciamento não garantindo qualidade às seqüências geradas, portanto, foi necessária a pesquisa de iniciadores que fossem capazes de seqüenciar com qualidade o gene *env* (**Figura 11**).

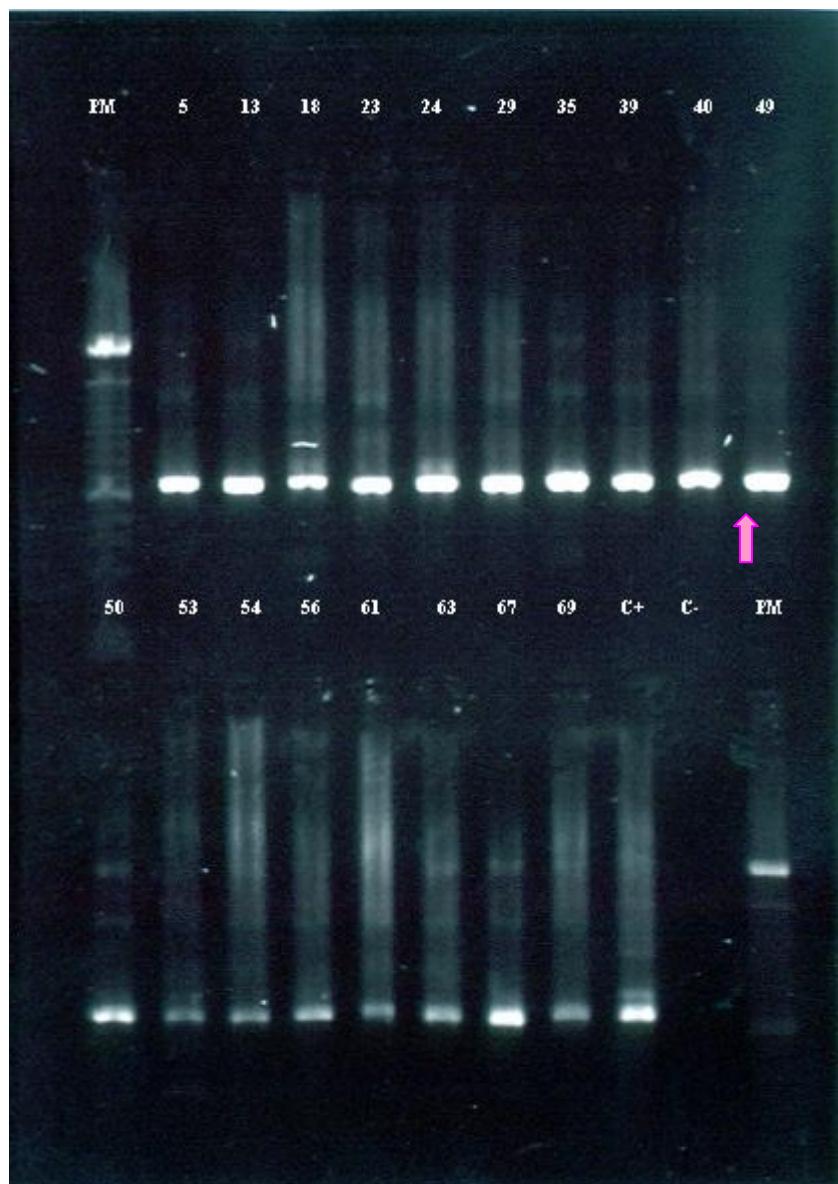
Assim, foram usados iniciadores mais internos no seqüenciamento do que os pesquisados e usados na amplificação, o que garantiu resultados ótimos. A **Figura 11** mostra algumas amostras no gel como exemplo de boa amplificação. Dessa forma, a metodologia final, que produz melhores bandas detectadas após exaustivas pesquisas é a descrita anteriormente na metodologia para o gene *env*, na qual utilizaram-se os iniciadores ED5/ED12 (*NIH aids Research and Reference Reagent Program Catalog, 2004-2005*) na primeira etapa da PCR-*nested*, enquanto os iniciadores ES7/ES8 (*NIH aids Research and Reference Reagent Program Catalog, 2004-2005*) foram utilizados na segunda etapa *nested*-PCR.

Cabe ressaltar que outros iniciadores como V3 *outer* e V3 *inner* também foram testados, mas como os resultados não mostraram amplificações as eletroforeses não foram fotografadas.





**Figura 10.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença, porém reduzida de bandas inespecíficas, mas com forte intensidade (seta azul) e bandas alvo não muito intensas (seta rosa). PM significa marcador de peso molecular de 100bp. Nas três fileiras do gel encontram-se identificadas as amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo.



**Figura 11.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença em alta intensidade da banda alvo (seta rosa) e a ausência de bandas inespecíficas. PM significa marcador de peso molecular de 100bp. Nas duas fileiras do gel encontram-se identificadas algumas amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo.

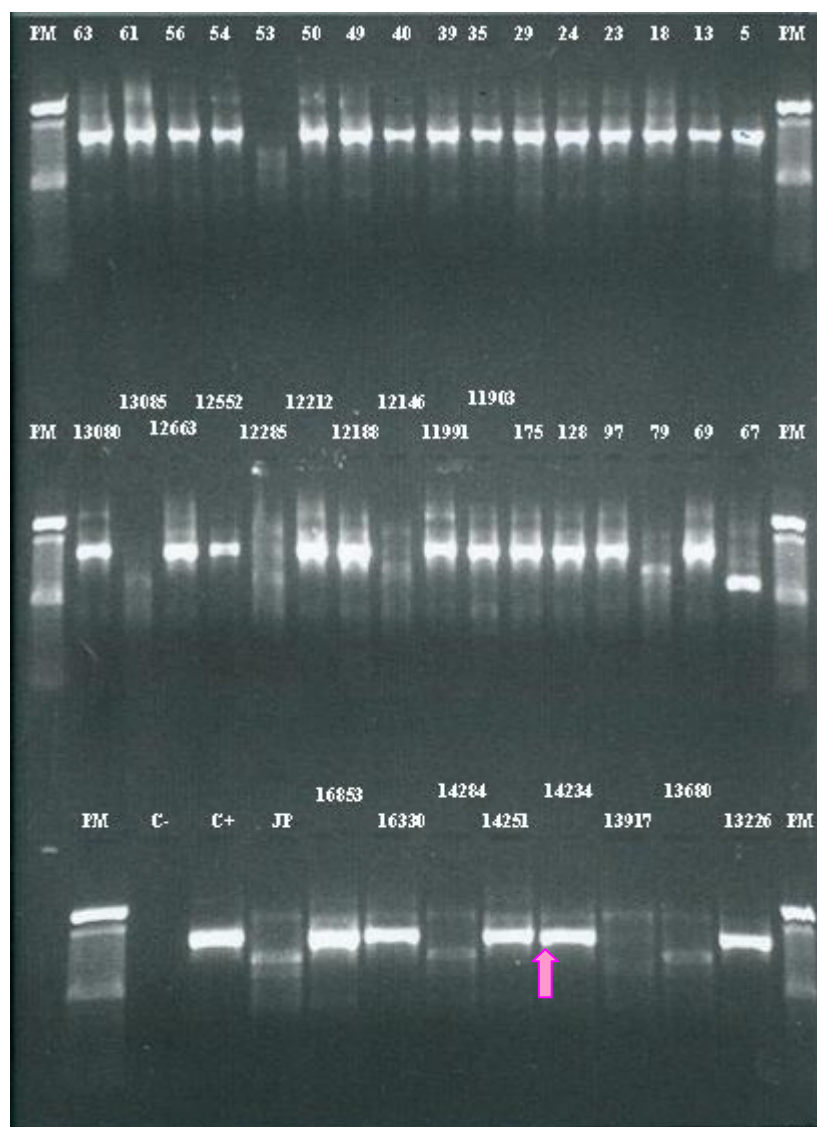
O produto obtido dessa amplificação foi do tamanho do esperado (655bp) (**Figura 11**), não apresentando inespecificidades. Como já dito, a **Figura 11** mostra apenas algumas amostras no gel como exemplo de boa amplificação. As amostras amplificadas foram então seqüenciadas para verificar se os iniciadores escolhidos flanqueavam a região de interesse do estudo e garantiam qualidade às seqüências. Com o resultado positivo dessa nova padronização, os produtos foram purificados, seqüenciados e analisados.

A partir desse resultado foram seqüenciadas 65 amostras para o gene *env* (GWGR) e 35 amostras para o gene *env* (GPGR). Das 65 amostras seqüenciadas, 5 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 60 amostras para GWGR. Das 35 amostras para GPGR seqüenciadas, 3 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 32 amostras para GPGR.

### 5.2.2 – GENE *pol*

O produto amplificado na segunda etapa da *nested*-PCR do gene *pol* do HIV-1 foi visualizado pela separação eletroforética em gel de agarose, onde as bandas apresentavam o tamanho esperado (1200 bp). A **Figura 12** mostra algumas amostras no gel como exemplo de boa amplificação. Embora o produto amplificado tenha demonstrado algumas bandas inespecíficas fracas, o protocolo utilizado para a amplificação desse gene está bem estabelecido no procedimento operacional padrão do Laboratório de Retrovirologia da UNIFESP e essas bandas inespecíficas não apresentaram nenhum problema na geração de seqüências, na edição e na análise do seqüenciamento desse gene.

A partir desse resultado foram seqüenciadas 65 amostras para o gene *pol* (GWGR) e 35 amostras para o gene *pol* (GPGR). Das 65 amostras seqüenciadas, 5 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 60 amostras para GWGR. Das 35 amostras para GPGR seqüenciadas, 3 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 32 amostras para GPGR.



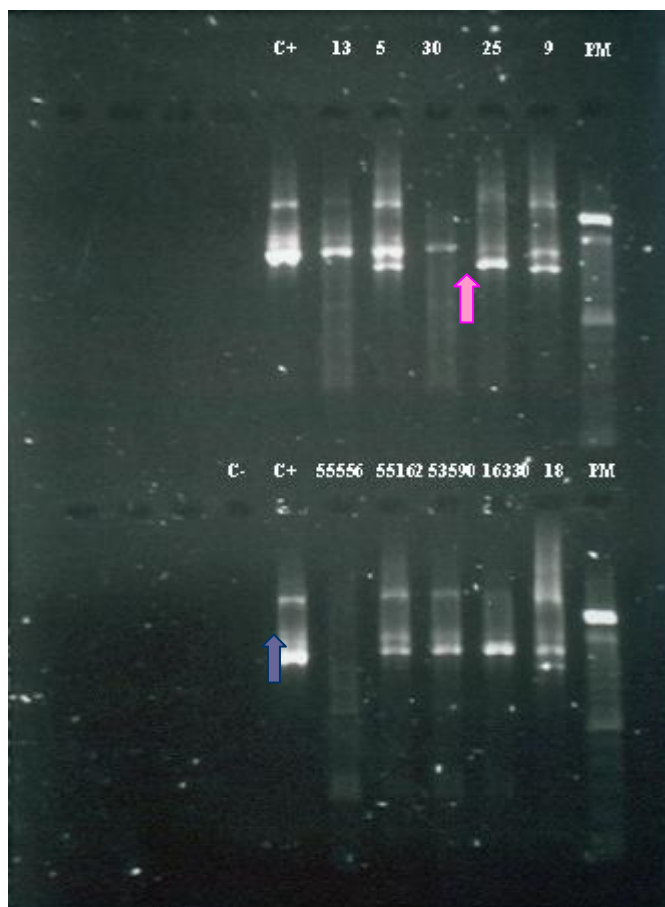
**Figura 12.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença em alta intensidade da banda alvo (seta rosa). PM significa marcador de peso molecular de 100bp. Nas três fileiras do gel encontram-se identificadas algumas amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo.

### 5.2.3 – GENE *gag*

O produto amplificado na segunda etapa da *nested*-PCR do gene *gag* do HIV-1 foi visualizado pela separação eletroforética em gel de agarose, onde as bandas apresentavam o tamanho esperado (1372 bp). A **Figura 13** mostra algumas amostras no gel como exemplo de boa amplificação. Contudo, apesar do produto amplificado apresentar algumas bandas inespecíficas, foi seqüenciado a fim de verificar se os iniciadores flanqueiam a região de interesse do estudo. Os

resultados obtidos confirmam a região estudada. A partir desses resultados analisamos se essas bandas inespecíficas não apresentavam nenhum problema na geração e na qualidade de seqüências, na edição e na análise das seqüências obtidas desse gene. Constatamos então que essas bandas inespecíficas não apresentavam nenhum problema na geração de seqüências, na edição e na análise do seqüenciamento desse gene.

A partir desse resultado foram seqüenciadas 65 amostras para o gene *gag* (GWGR) e 35 amostras para o gene *gag* (GPGR). Das 65 amostras seqüenciadas, 5 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 60 amostras para GWGR. Das 35 amostras para GPGR seqüenciadas, 3 não apresentaram bom seqüenciamento, não permanecendo no estudo, o que resultou num número de 32 amostras para GPGR.



**Figura 13.** Eletroforese em gel de agarose 1,5% do produto amplificado na segunda etapa da *nested*-PCR, onde verifica-se a presença da banda alvo (seta rosa) e inespecificidades (seta azul). PM significa marcador de peso molecular de 100bp. Nas duas fileiras do gel encontram-se identificadas algumas amostras utilizadas no teste. C+ é controle positivo e C- é controle negativo.

### 5.3 – ANÁLISE DOS SUBTIPOS E DETECÇÃO DE RECOMBINAÇÃO

Após o alinhamento, as seqüências selecionadas foram subtipadas por meio de duas metodologias, para que nenhuma amostra pertencente a outros subtipos não B fosse analisada (**Tabela 13**).

Uma metodologia empregada foi o SIMPLOT (Salminen *et al*, 1995) disponível no programa SIMPLOT versão 3.5.1. As **Figuras 14 e 15** mostram exemplos de amostras submetidas ao SIMPLOT subtipadas como B e as **Figuras 16 e 17** mostram exemplos de amostras subtipadas como não B. Nessa análise foram utilizadas janelas de 250 bp com incremento de 10bp para o gene *env*, janelas de 600 bp com incremento de 200 bp para o gene *gag* e janelas de 400 bp com incremento de 60 bp para o gene *pol*. A janela corresponde ao que o método lê no momento. Elas são determinadas de acordo com o tamanho do fragmento e quão variável ele é. Dessa forma, para fragmentos menores e mais variáveis, como o *env*, por exemplo, a janela deve ser menor. O contrário é verdadeiro, por isso, as janelas aumentaram de acordo com o tamanho do fragmento do gene estudado.

As seqüências selecionadas também foram subtipadas utilizando-se o método de *Maximum Likelihood* (máxima verossimilhança) implementado pelo *software* PAUP\* e o modelo evolutivo apontado para cada árvore de cada gene foi GTR+G (Swofford, 2002). Posteriormente, com intuito de análise filogenética adequada, como já explicado anteriormente, as árvores foram construídas utilizando-se apenas amostras subtipo B (**Figuras 18 a 20**).

Essas análises mostraram que dentre as 60 seqüências dos genes *env*, *gag* e *pol* para GWGR, 37 pertencem ao subtipo B, 18 apresentam recobinação BF em algum dos genes estudados, 3 apresentam recobinação BC em algum dos genes estudados, 1 apresenta subtipo C em algum dos genes estudados e 1 apresenta recombinante CPX\_18 em algum dos genes estudados (**Tabela 11**).

Nas 32 amostras dos genes *env*, *gag* e *pol* para GPGR, 29 pertencem ao subtipo B e 3 apresentam recobinação BF em algum dos genes estudados (**Tabela 12**).

Sendo assim, com a exclusão das amostras não B, o número de amostras GWGR analisadas e consideradas para o estudo mudou para 37 (**Tabela 11**) e o número de amostras GPGR analisadas e consideradas para o estudo mudou de 32 para 29 (**Tabela 12**).

**Tabela 11:** Número de amostras GWGR subtipadas e utilizadas no estudo

Amostras GWGR						
Número de amostras iniciais	Amostras subtipo B	Amostras subtipo C	Amostras BF	Amostras BC	CRF_CPX_18	Número final de amostras (apenas B)
60	37	1	18	3	1	37

**Tabela 12:** Número de amostras GPGR subtipadas e utilizadas no estudo

Amostras GPGR			
Número de amostras iniciais	Amostras subtipo B	Amostras BF	Número final de amostras (apenas B)
32	29	3	29

A relação completa das amostras GWGR e GPGR subtipadas para seus respectivos genes pode ser observada na **Tabela 13**.

Como já dito, as seqüências selecionadas também foram subtipadas utilizando-se o método de *Maximum Likelihood* (máxima verossimilhança) implementado pelo *software* PAUP\* (Swofford, 2002) (**Figuras 18 a 20**).

A árvore contruída para o gene *env* (**Figura 18**) mostrou que as seqüências que contém W no *loop* V3 não formam nenhum agrupamento monofilético dentro do subtipo B, as seqüências independentemente de possuírem o aminoácido W ou P distribuem-se aleatoriamente na árvore de *env*.

A árvore construída para o gene *pol* com as variantes GWGR e GPGR (**Figura 19**) revelou que as amostras se agrupam no subtipo B e ambas as variantes GWGR e GPGR se misturam. No entanto, essa árvore demonstrou a presença de dois grupos de amostras, porém, pode-se visualizar tanto amostras GWGR e GPGR misturadas em ambos os grupos de *pol*.

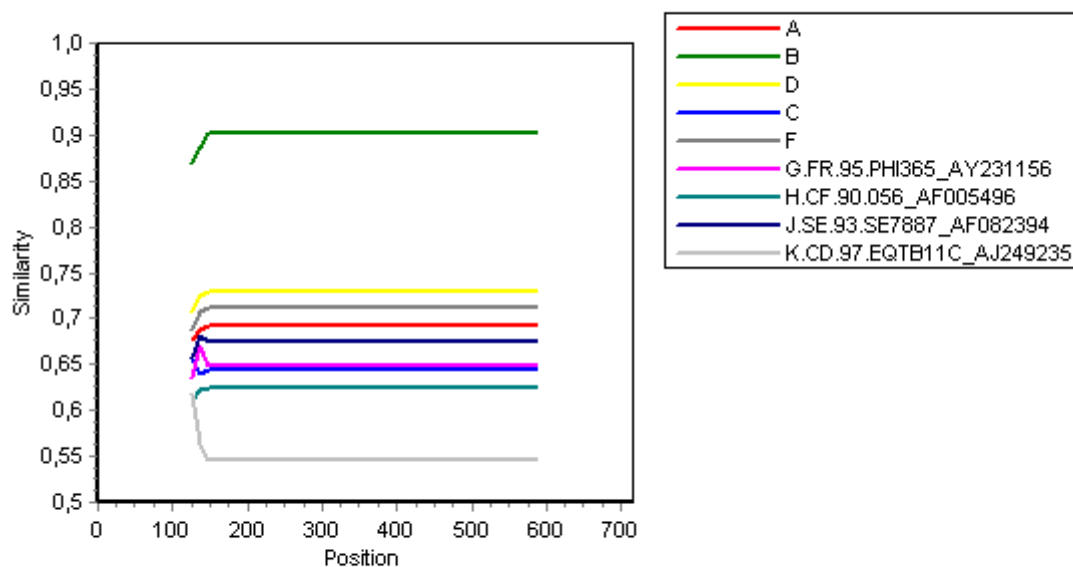
A inferência filogenética da região *gag* (**Figura 20**) também demonstrou que as amostras GWGR e GPGR não diferenciam-se nessa região.

Para realmente definir se amostras GWGR e GPGR são variantes distintas ou amostras que contém apenas uma assinatura diferenciada em *env*, foi composta uma seqüência concatêmera, onde foi comparada a distribuição de ambas as amostras em um agrupamento filogenético. Para tal, foram unidas as regiões *gag*, *pol* e *env* de cada amostra GWGR e GPGR, compondo dessa forma um “mini-genoma” de quase 3.000 nucleotídeos, o que dá uma maior robustez para a distribuição encontrada. A distribuição encontrada demonstrou que não houve separação das amostras GWGR e GPGR e sim uma total mistura entre elas (**Figura 21**).

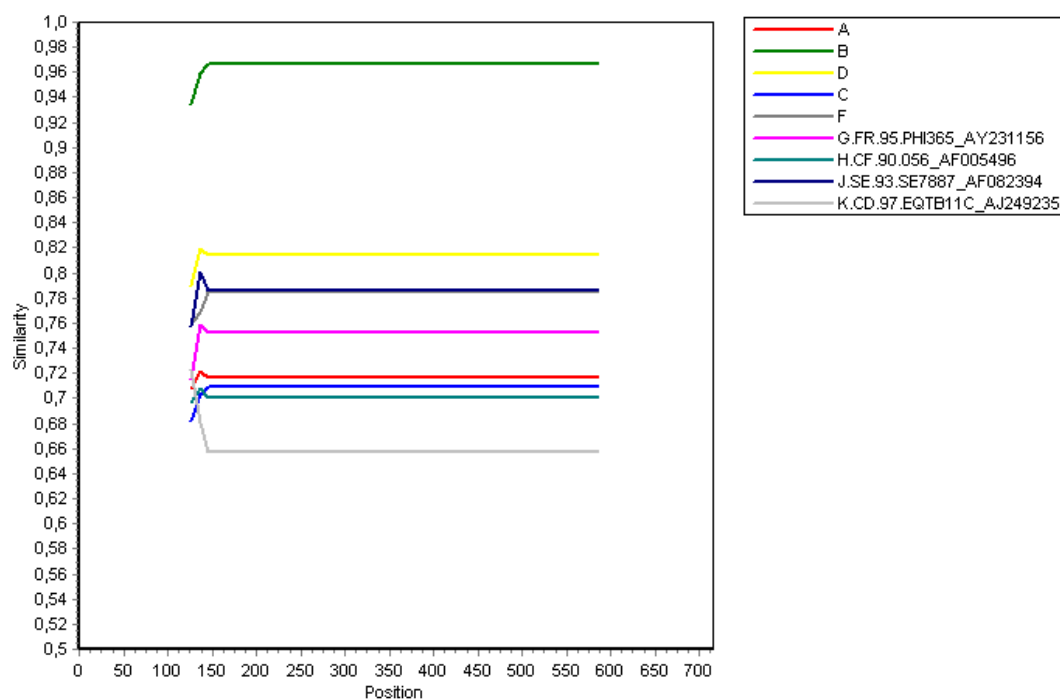
**Tabela 13:** Identificação das amostras e seus subtipos

<b>Amostras <i>gw</i></b>	<b>Subtipo <i>env gw</i></b>	<b>Subtipo <i>pol gw</i></b>	<b>Subtipo <i>gag gw</i></b>	<b>Amostras <i>gp</i></b>	<b>Subtipo <i>env gp</i></b>	<b>Subtipo <i>pol gp</i></b>	<b>Subtipo <i>gag gp</i></b>
5	B	BF	BF	1A	B	B	B
13	B	B	B	4A	B	BF	B
18	B	BF	BF	5A	B	B	B
23	B	BF	BF	6A	B	B	B
24	B	B	B	8A	BF	B	B
29	B	B	B	9A	B	B	B
31	BF	B	B	10A	B	B	B
35	B	BF	BF	12A	B	B	B
39	B	B	B	14A	B	B	B
40	B	B	BF	15A	B	B	B
41	BF	B	B	19A	B	B	B
49	B	B	B	21A	B	B	B
50	B	B	B	22A	B	B	B
53	B	B	B	23A	B	B	B
54	B	B	B	24A	B	B	B
56	B	B	B	26A	BF	B	B
61	B	B	B	27A	B	B	B
63	B	B	B	28A	B	B	B
67	B	B	B	34A	B	B	B
69	B	B	B	35A	B	B	B
79	B	B	B	36A	B	B	B
87	B	B	B	41A	B	B	B
97	B	BF	BF	43A	B	B	B
128	B	B	B	44A	B	B	B
132	B	B	BC	46A	B	B	B
133	B	B	B	47A	B	B	B
134	B	BF	B	48A	B	B	B
135	B	B	B	50A	B	B	B
143	B	B	B	51A	B	B	B
175	B	B	BF	52A	B	B	B
181	B	B	B	53A	B	B	B
188	B	B	B	56A	B	B	B
190	B	B	B				
11903	B	B	B				
11991	B	B	B				
12188	B	B	CRF_CPX_18				
12212	B	B	B				
12285	B	BF	B				
12552	B	B	B				
12663	B	B	C				
13080	B	BF	BF				
13085	B	B	B				
13680	B	BF	B				
13226	B	BF	BF				
14234	B	B	BF				
14251	B	B	BC				
16330	B	BF	BF				
16853	B	B	BF				
53590	B	B	B				
55556	BF	B	B				
58270	B	B	B				
59381	B	B	B				
61450	B	BC	C				
62023	B	B	B				
64054	B	B	B				
64230	B	B	B				
64445	B	B	B				
65822	B	BF	B				
66116	B	B	B				
66666	B	B	B				

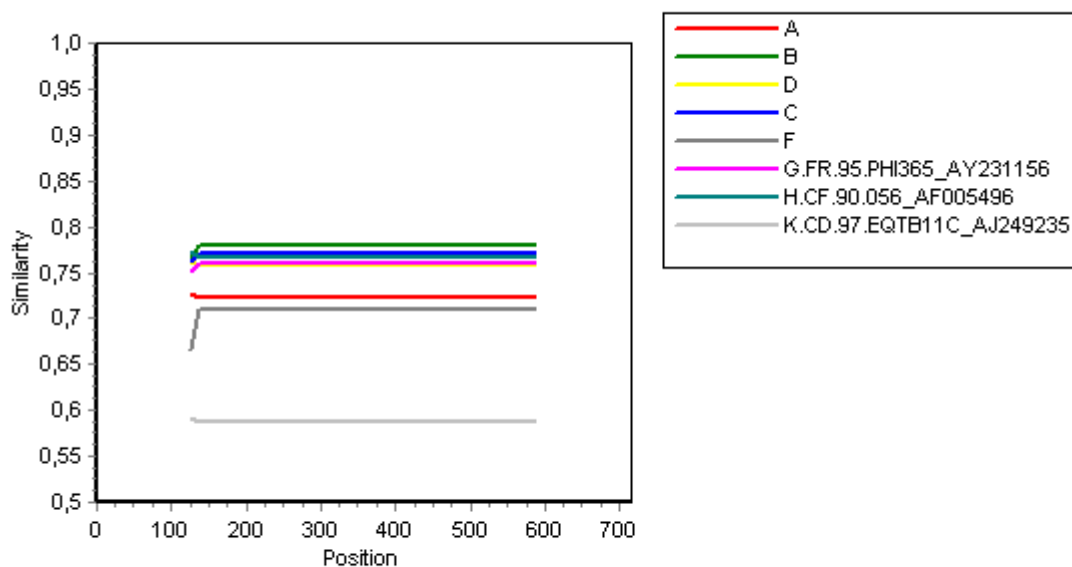




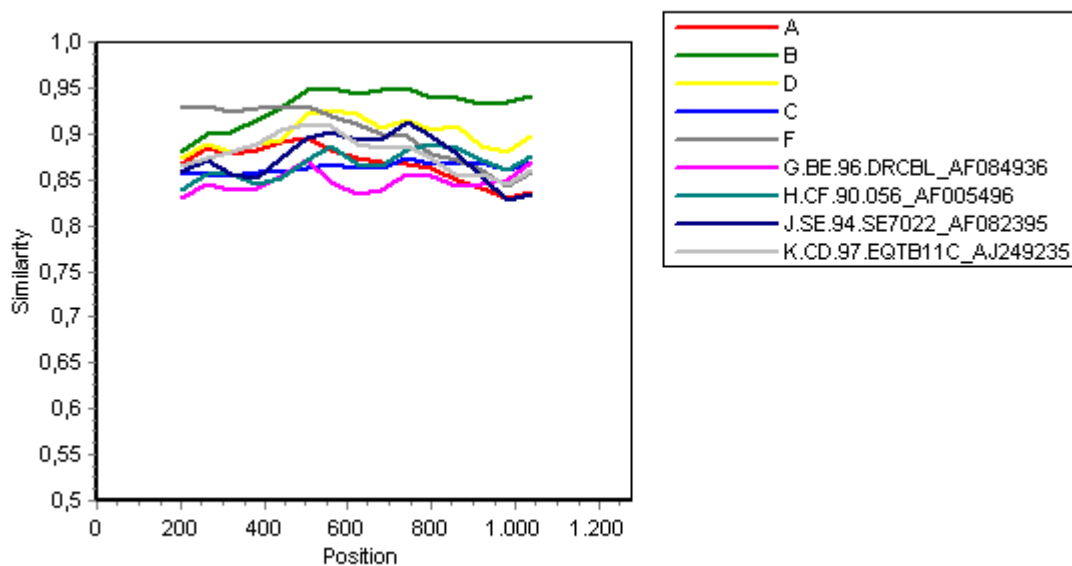
**Figura 14:** Exemplo de amostra subtipo B segundo a metodologia do SIMPLOT



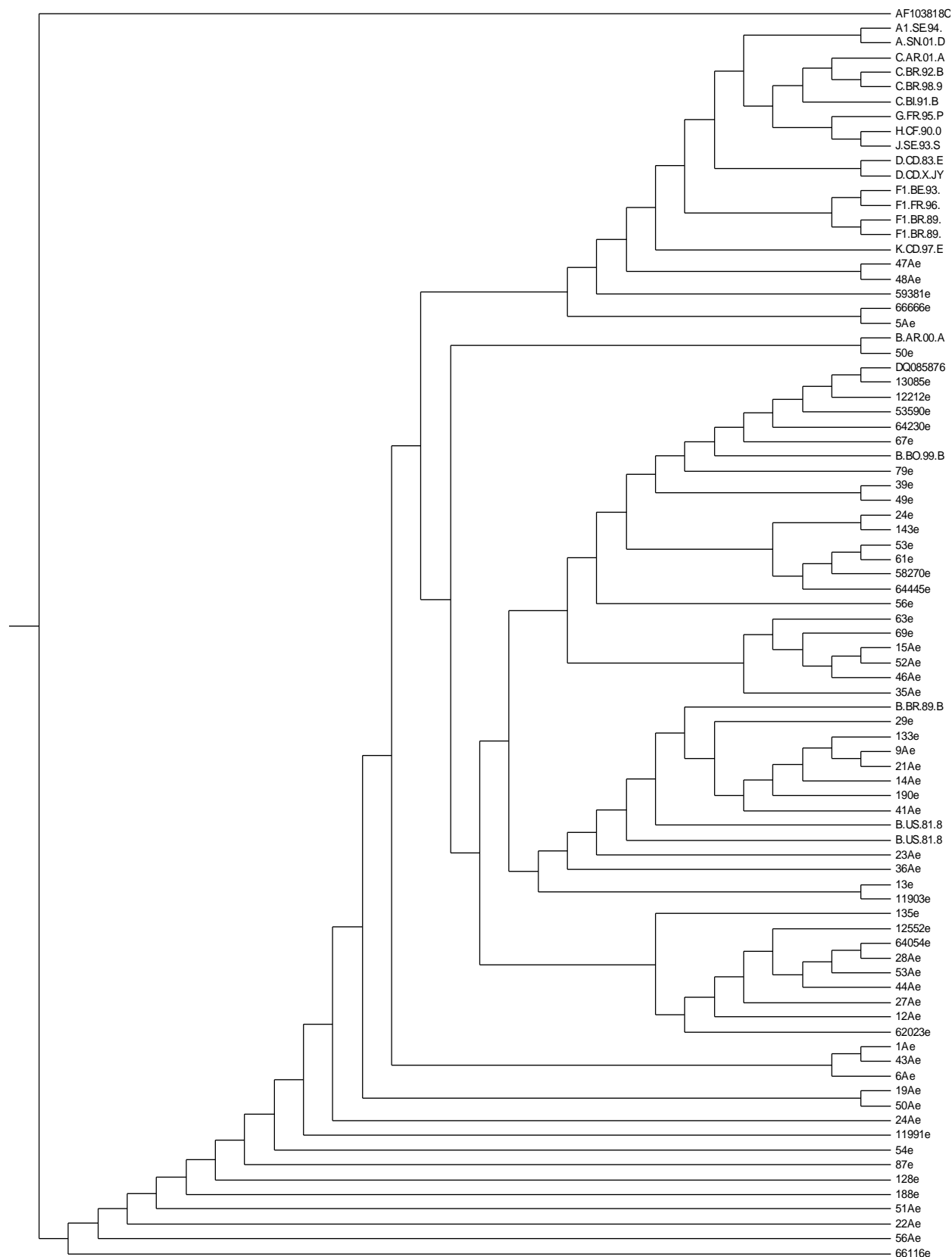
**Figura 15:** Exemplo de amostra subtipo B segundo a metodologia do SIMPLOT



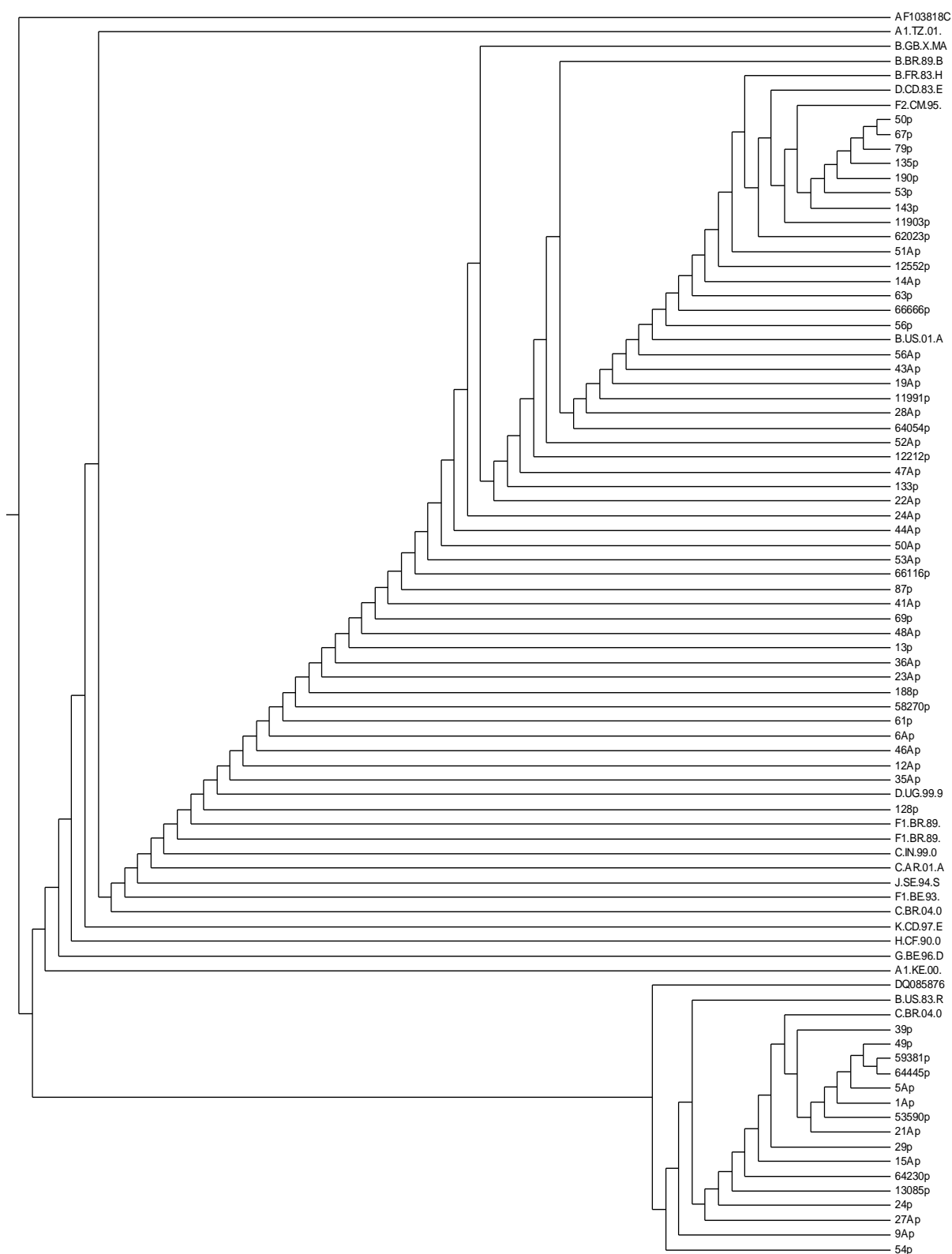
**Figura 16:** Exemplo de amostra subtipo não B segundo a metodologia do SIMPLOT



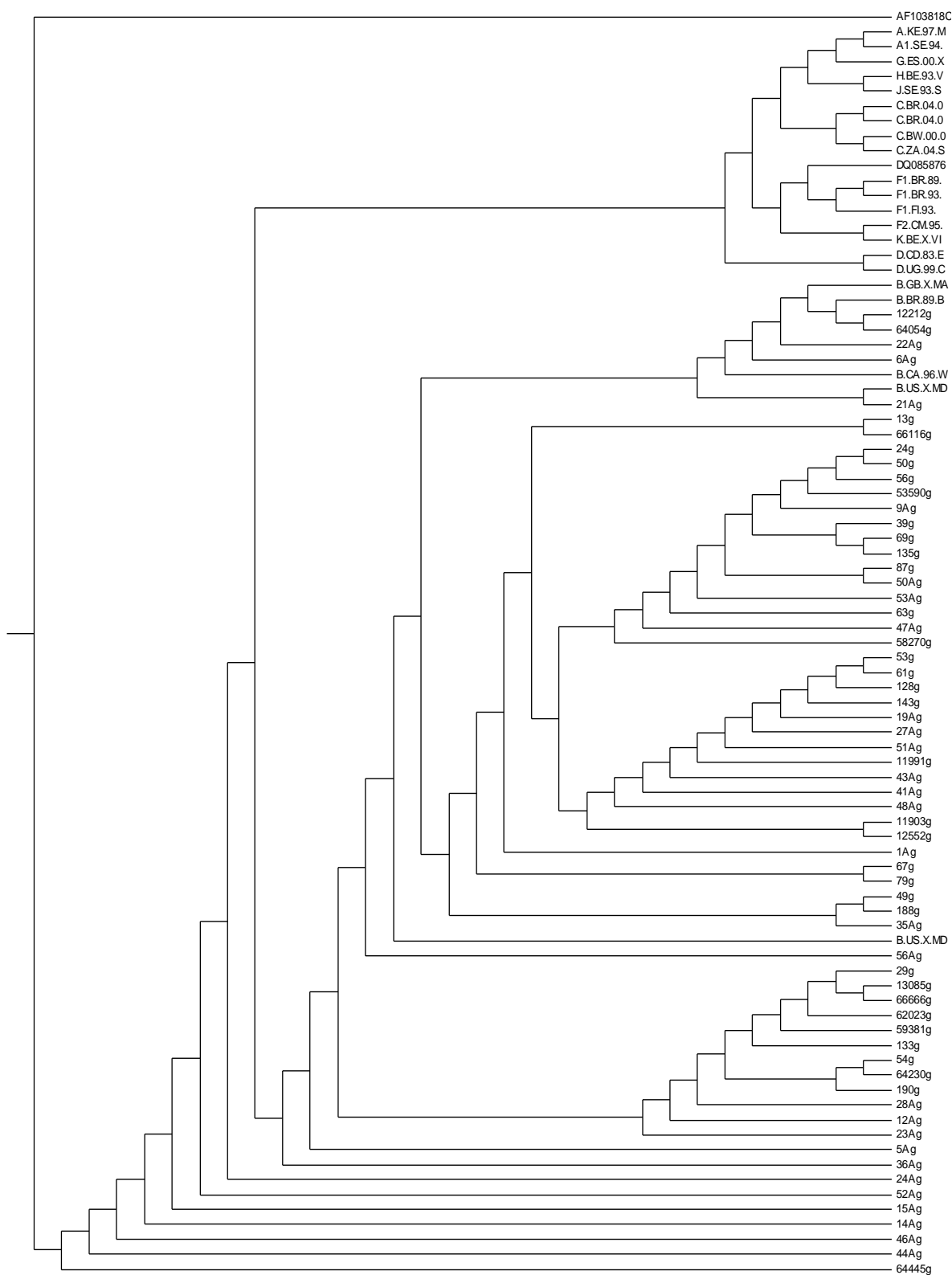
**Figura 17:** Exemplo de amostra subtipo não B segundo a metodologia do SIMPLOT



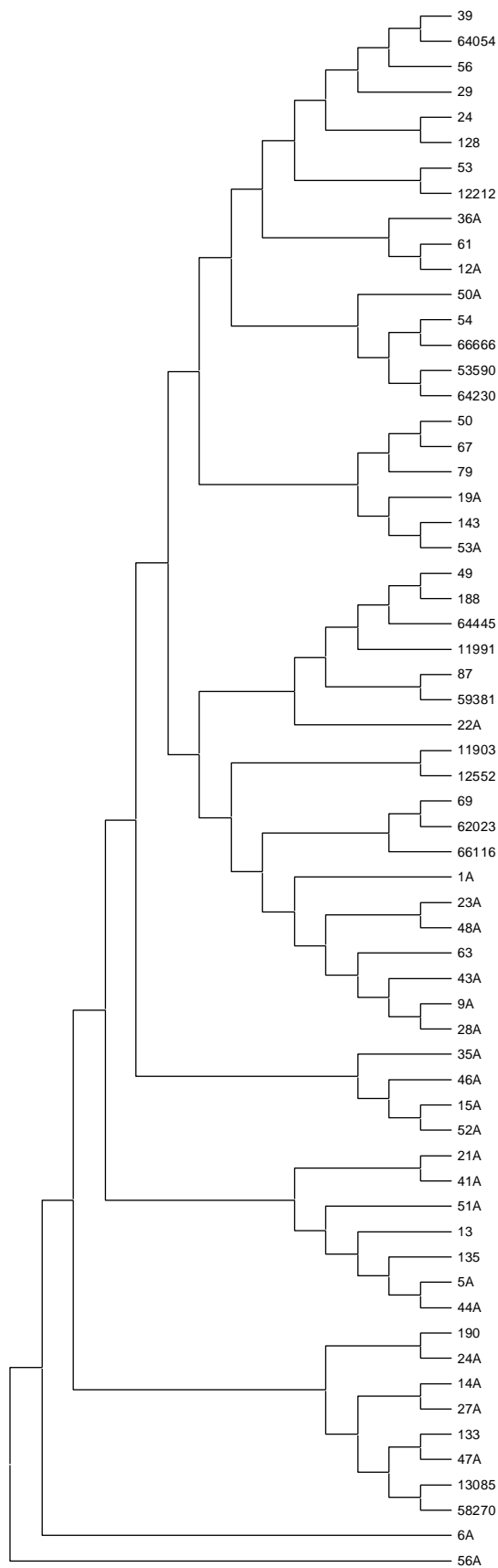
**Figura 18:** Árvore construída por *Maximum Likelihood* através do software PAUP, tendo como modelo evolutivo apontao o GTR+G. Amostras do gene *env* com as variantes GWGR e GPGR. As amostras GWGR apresentam a letra "e" após seu número e as amostras GPGR apresentam as letras "Ae" após seu número. As demais seqüências são referências.



**Figura 19:** Árvore construída por *Maximum Likelihood* através do software PAUP, tendo como modelo evolutivo apontado o GTR+G. Amostras do gene *pol* com as variantes GWGR e GPGR. As amostras GWGR apresentam a letra "p" após seu número e as amostras GPGR apresentam as letras "Ap" após seu número. As demais seqüências são referências.



**Figura 20:** Árvore construída por *Maximum Likelihood* através do software PAUP, tendo como modelo evolutivo apontado o GTR+G. Amostras do gene *gag* com as variantes GWGR e GPGR. As amostras GWGR apresentam a letra "g" após seu número e as amostras GPGR apresentam as letras "Ag" após seu número. As demais seqüências são referências.



**Figura 21:** Árvore construída por *Maximum Likelihood* através do software PAUP, tendo como modelo evolutivo o GTR+G. Amostras dos genes *env*, *pol* e *gag* com as variantes GWGR e GPGR formando um concatâmero. As amostras GWGR apresentam apenas números e as amostras GPGR apresentam as letras "A" após seu número.

#### 5.4 – ANÁLISE DA PRESSÃO SELETIVA

Após a determinação do genótipo das amostras, passa-se à análise da pressão seletiva nas seqüências do HIV-1 subtipo B. Dessa forma, foram excluídas seqüências com *indels* e/ou ambigüidades. As remanescentes foram separadas em seis grupos, de acordo com o gene e a variante: *env* (GWGR), *pol* (GWGR), *gag* (GWGR), *env* (GPGR), *pol* (GPGR) e *gag* (GPGR).

Inicialmente, a abordagem foi feita através do método de Nei & Gojobori para estimar o número de substituições não sinônimas ( $dN$ ) e sinônimas ( $dS$ ), códon a códon (**Tabela 14** e **Figura 22**). De acordo com as análises de pressão seletiva baseadas em mensuração par a par (Nei & Gojobori) encontramos uma pressão seletiva com valores de  $\omega$  próximos de 1,5 na região do *env* tanto para amostras GWGR como GPGR. Na região *gag* foram encontrados valores de  $\omega$  também próximos de 1,5 em amostras GWGR e GPGR. Na região *pol* as amostras GWGR apresentaram um valor de  $\omega$  também similar a 1,5, porém as amostras GPGR mostraram valores de  $\omega$  acima de 2.

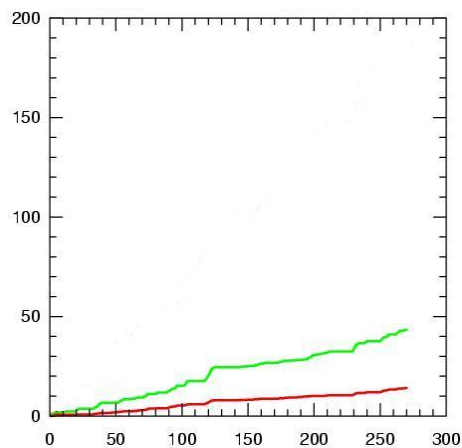
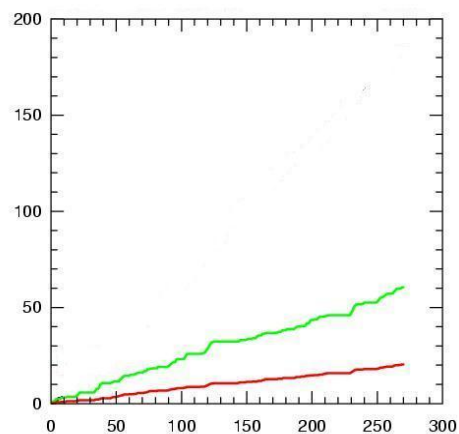
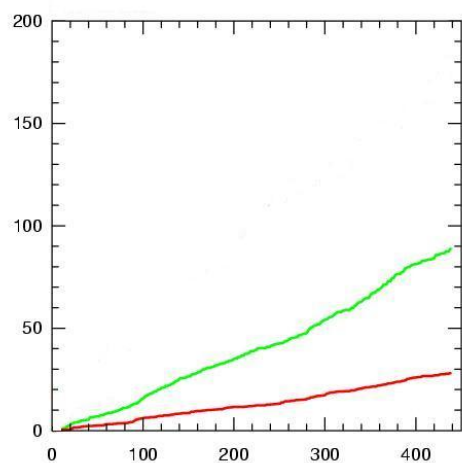
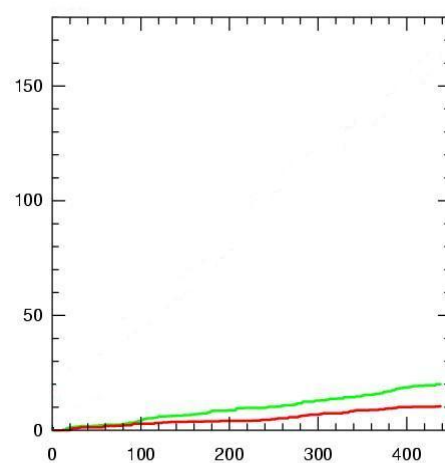
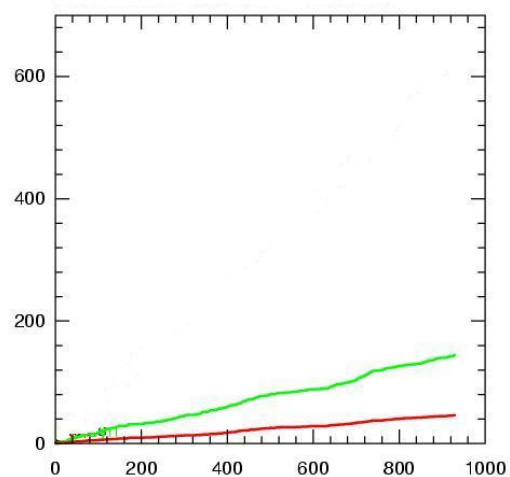
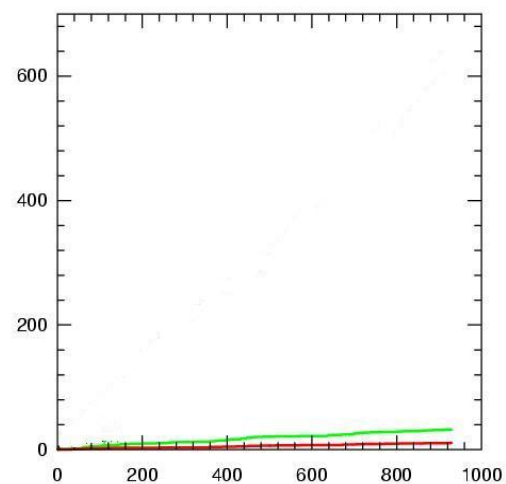
Para o cálculo da diversidade genética (**Tabela 15**), como mencionado anteriormente, as medidas de diversidade foram baseadas em comparação par a par das distâncias entre as seqüências usando-se a matriz de substituição Kimura 2 Parâmetros, que dá um peso duas vezes maior para as transversões em relação às transições. Para tal análise também separamos as seqüências de acordo com as regiões gênicas e com as variantes GWGR e GPGR. Nas amostras GPGR foi encontrada uma diversidade de 0,36 na região do *env* e de 0,01 e 0,02 nas regiões *gag* e *pol*, respectivamente. Nas amostras GWGR encontramos uma diversidade de 0,18 no *env* e 0,22 e 0,26 nas regiões *gag* e *pol*, respectivamente.

Apesar das análises feitas individualmente com as seqüências dos grupos indicarem similaridades na posição dos códons sob seleção positiva em cada gene e no montante global da razão  $dN/dS$ , essas análises não exploraram formalmente se existem diferenças códon-por-códon entre as populações distintas. Em razão disso, um teste foi feito para determinar se existem códons em cada um dos seis grupos sob pressões distintas de seleção. Esse teste usa filogenias e indica (baseado no LRT e no  $p$ -value) quais códons estão diferentemente selecionados entre os grupos.

O resultados desse teste baseado no LRT e no  $p$ -value mostrou que a distribuição dos códons sob seleção positiva em cada gene é muito parecida. Assim, códons sob seleção em *env* GWGR são parecidos com os códons sob seleção em *env* GPGR. O mesmo ocorre com *pol* e *gag* (**Tabela 16**). Os resultados obtidos com essa análise indicaram códons sob seleção positiva ao longo de toda a seqüência da região *env* do HIV-1 em ambas as variantes. Cabe destacar que códons presentes na coroa do *env*, resíduos 306 e 320 apresentaram pressão seletiva em ambas

variantes GWGR e GPGR (**Tabela 16**). O mesmo modelo evolutivo, o M2, foi escolhido para ambos conjuntos de variantes, sendo os valores de  $\omega$  bem parecidos, em torno de 1,2 para amostras GWGR e 1,3 para amostras GPGR. Foram detectados 7 códons sob pressão seletiva em amostras GWGR e GPGR. Em *gag* foi detectada uma menor quantidade de códons sob pressão positiva em ambas as variantes. Tanto em *gag* de amostras GWGR como GPGR os códons estão sob pressão positiva e em ambas as variantes os valores de  $\omega$  em *gag* foram de 1,4. O modelo selecionado foi o M2. Na região *pol* foram identificados muitos códons sob seleção positiva. Nas amostras GWGR foram identificados 12 códons sob pressão seletiva sendo a grande maioria códons de resistência aos anti-retrovirais. Nas amostras GPGR foram identificados 9 códons sob pressão seletiva também correspondendo a códons de resistência anti-retroviral. Os valores de  $\omega$  foram de 1,40 para amostras GWGR e 2,26 para amostras GPGR, sendo o modelo evolutivo escolhido, M2, o mesmo para ambas as variantes.



***env* GWGR*****env* GPGR*****pol* GWGR*****pol* GPGR*****gag* GWGR*****gag* GPGR**

**Figura 22:** Média do comportamento das mutações sinônimas e não sinônimas em toda a região codificada. O eixo das ordenadas representa a porcentagem e o eixo das abscissas representa o tamanho da região codificada. A cor verde representa dN e a cor vermelha representa dS.

**Tabela14:** Valores de substituições sinônimas e não sinônimas por *Nei & Gojobori*

<i>Gene env</i>			
Variante	dN	dS	$\omega$ (dN/dS)
<b>GWGR</b>	0,3258	0,2752	1,1896
<b>GPGR</b>	0,5740	0,4228	1,3075
<i>Gene pol</i>			
Variante	dN	dS	$\omega$ (dN/dS)
<b>GWGR</b>	0,3069	0,2123	1,4015
<b>GPGR</b>	0,0730	0,0339	2,2608
<i>Gene gag</i>			
Variante	dN	dS	$\omega$ (dN/dS)
<b>GWGR</b>	0,3589	0,2658	1,4101
<b>GPGR</b>	0,0740	0,0527	1,4209

$\omega > 1$  (dN>dS ) para seleção positiva;

$\omega < 1$  (dN<dS ) para seleção negativa ou purificadora;

$\omega = 1$  (dN=dS ) para um regime neutro de substituições.

**Tabela15:** Valores de Diversidade Genética e Desvio Padrão

<i>Gene env</i>		
Variante	Diversidade Genética	Desvio Padrão
<b>GWGR</b>	0,184	0,013
<b>GPGR</b>	0,368	0,021
<i>Gene pol</i>		
Variante	Diversidade Genética	Desvio Padrão
<b>GWGR</b>	0,229	0,008
<b>GPGR</b>	0,013	0,002
<i>Gene gag</i>		
Variante	Diversidade Genética	Desvio Padrão
<b>GWGR</b>	0,267	0,029
<b>GPGR</b>	0,021	0,002

**Tabela16:** Análise por *Maximum Likelihood* dos genes *env*, *pol* e *gag* do HIV-1 para as amostras GWGR e GPGR

Modelo	$-n\ln L$	Parâmetros Estimados	Seleção Positiva códon com $p > 0.999$
<b>Gene <i>env</i> (GWGR)</b>			
M0 ( <i>one ratio</i> )	3127.32011	$\omega_0 = 0.4327$ $\kappa$ (ts/tv) = 2.11437	<i>Não permite</i>
		$p_0 = 0.65225$ $p_1 = 0.34775$	
M1 ( <i>nearlyneutral</i> )	3018.35208	$\omega_0 = 0.00000$ $\omega_1 = 1.00000$	<i>Não permite</i>
		$\kappa$ (ts/tv) = 2.01824	
		$p_0 = 0.63301$ $p_1 = 0.30331$	
		$p_2 = 0.0637$	
M2 ( <i>selection</i> ) *	3004.46263 <sup>1</sup>	$\omega_0 = 0.04273$ $\omega_1 = 1.000$ $\omega_2 =$ 1.1896	33, 45, 59, 64, 65, 306, 320
		$\kappa$ (ts/tv) = 2.59750	
		$p_0 = 0.60233$ $p_1 = 0.3184$	
		$p_2 = 0.07928$	
M3 ( <i>discrete</i> )	3003.98168	$\omega_0 = 0.03146$ $\omega_1 = 0.8169$ $\omega_2 =$ 1.1723	33, 45, 64, 65, 306, 320
		$\kappa$ (ts/tv) = 2.25216	
<b>Gene <i>env</i> (GPGR)</b>			
M0 ( <i>one ratio</i> )	3351.23052	$\omega_0 = 0.4067$ $\kappa$ (ts/tv) = 2.92624	<i>Não permite</i>
		$p_0 = 0.64067$ $p_1 = 0.35933$	
M1 ( <i>nearlyneutral</i> )	3253.6001	$\omega_0 = 0.00000$ $\omega_1 = 1.00000$	<i>Não permite</i>
		$\kappa$ (ts/tv) = 2.48571	
		$p_0 = 0.6270$ $p_1 = 0.31451$ $p_2 =$ 0.0584	
M2 ( <i>selection</i> ) *	3244.95198 <sup>2</sup>	$\omega_0 = 0.0688$ $\omega_1 = 1.000$ $\omega_2 = 1.3075$	33, 45, 59, 64, 65, 306, 320
		$\kappa$ (ts/tv) = 2.67378	

		$p_0=0.5582$	$p_1=0.3452$	$p_2=0.0965$	
M3	3242.81107	$\omega_0=0.0420$	$\omega_1=0.6729$	$\omega_2=1.2865$	33, 45, 65, 306, 320
(discrete)		$\kappa$ (ts/tv)= 2.61661			
<b>Gene pol (GWGR)</b>					
M0 (one ratio)	3127.32011	$\omega_0=0.4327$	$\kappa$ (ts/tv)= 2.11437		Não permite
		$p_0=0.65225$ $p_1=0.34775$			
M1	3018.35208	$\omega_0=0.00000$	$\omega_1=1.00000$		Não permite
(nearlyneutral)		$\kappa$ (ts/tv)= 2.01824			
		$p_0=0.63301$ $p_1=0.30331$			
		$p_2=0.0637$			
M2	3004.46263 <sup>1</sup>	$\omega_0=0.04273$	$\omega_1=1.000$	$\omega_2=$	10, 14, 39, 41, 45, 59, 62, 67, 70, 101, 210, 184
(selection) *		1.4015			
		$\kappa$ (ts/tv)= 2.59750			
		$p_0=0.60233$ $p_1=0.3184$			
		$p_2=0.07928$			
M3	3003.98168	$\omega_0=0.03146$	$\omega_1=0.8169$	$\omega_2=$	10, 14, 39, 41, 45, 59, 62, 67, 70, 101, 210,
(discrete)		1.3824			215, 184
		$\kappa$ (ts/tv)= 2.25216			
<b>Gene pol (GPGR)</b>					
M0 (one ratio)	5513.034008	$\omega_0=0.5139$	$\kappa$ (ts/tv)= 2.34002		Não permite
		$p_0=0.65566$ $p_1=0.34434$			
M1	5258.007758	$\omega_0=0.07071$	$\omega_1=1.00000$		Não permite
(nearlyneutral)		$\kappa$ (ts/tv)= 2.14265			
		$p_0=0.63352$ $p_1=0.31150$ $p_2=$			
		0.0549			
M2	5200.17740 <sup>5</sup>	$\omega_0=0.07635$	$\omega_1=1.000$	$\omega_2=2.2608$	10, 14, 39, 41, 67, 70, 101, 210, 184
(selection) *					

		$\kappa$ (ts/tv)= 2.34630	
		$p_0= 0.64368$ $p_1= 0.30386$ $p_2=$ 0.0524	
M3 (discrete)	5199.472421	$\omega_0=0.08384$ $\omega_1=1.11840$ $\omega_2=$ 2.2498	14, 41, 45, 59, 62, 67, 70, 101, 210, 215, 184
		$\kappa$ (ts/tv)= 2.37786	
<b>Gene gag (GWGR)</b>			
M0 (one ratio)	10353.50682	$\omega_0=0.4401$ $\kappa$ (ts/tv)= 2.31434	<i>Não permite</i>
		$p_0= 0.67323$ $p_1= 0.32677$	
M1 (nearlyneutral)	9726.004577	$\omega_0=0.06092$ $\omega_1=1.00000$ $\kappa$ (ts/tv)= 2.19269	<i>Não permite</i>
		$p_0= 0.66561$ $p_1= 0.25875$ $p_2=$ 0.0756	
M2 (selection) *	9644.09090 <sup>4</sup>	$\omega_0=0.06974$ $\omega_1=1.000$ $\omega_2= 1.4101$ $\kappa$ (ts/tv)= 2.41124	34, 62, 98
		$p_0= 0.66264$ $p_1= 0.25354$ $p_2=$ 0.0838	
M3 (discrete)	9643.55413	$\omega_0=0.06727$ $\omega_1=0.92411$ $\omega_2=$ 1.4072 $\kappa$ (ts/tv)= 2.39518	15, 98
<b>Gene gag GPGR</b>			
M0 (one ratio)	8580.184763	$\omega_0=0.4202$ $\kappa$ (ts/tv)= 2.20087	<i>Nenhuma</i>
		$p_0= 0.69820$ $p_1= 0.30180$	
M1 (nearlyneutral)	8073.326331	$\omega_0=0.00000$ $\omega_1=1.00000$ $\kappa$ (ts/tv)= 2.11126	<i>Não encontrada</i>
		$p_0= 0.64267$ $p_1= 0.26784$ $p_2=$ 0.0895	
M2	7963.27659 <sup>2</sup>		34, 62, 98

(selection) *		$\omega_0=0.07113$	$\omega_1=1.000$	$\omega_2= 1.4209$	
		$\kappa$ (ts/tv)= 2.39112			
		$p_0= 0.59467$			
		$p_1= 0.30683$			
		$p_2=$			
		0.0985			
M3	7951.84511 <sup>3</sup>	$\omega_0=0.04524$	$\omega_1=0.67408$	$\omega_2=$	15, 76, 98
(discrete)		1.4177			
		$\kappa$ (ts/tv)= 2.27345			

---

\*) Melhor modelo indicado pelo teste LRT.  $p$ =proporção de códons pertencente a cada categoria de  $\omega$  ( $d_N/d_S$ ) definido pelos diferentes modelos.  $\kappa$  (ts/tv)= taxa de transição/transversão. 1)  $\Delta$ LRT M2vsM3 taxa=0.961914,  $p=0.618191$ . 2)  $\Delta$ LRT M2vsM3 taxa=4.281738  $p=0.117553$ . 3)  $\Delta$ LRT M2vsM3 taxa=22.862305  $p=0.000011$ , 4)  $\Delta$ LRT M2vsM3 taxa=1.074219  $p=0.584435$ . 5)  $\Delta$ LRT M2vsM3 taxa=1.409180  $p=0.494311$ .

## 6.0 - DISCUSSÃO

Apesar do subtipo B ser o mais predominante na epidemia de aids no Brasil, existem evidências que esse subtipo não é homogêneo. O subtipo B apresenta no Brasil uma assinatura no tetrâmero da coroa V3 da gp120 que é a presença de um triptofano (W) ao invés de prolina (P) que é o aminoácido mais freqüente nessa posição do genoma do HIV-1. Aproximadamente metade dos isolados do subtipo B no Brasil possuem um triptofano (W) no tetrâmero do *loop* V3 da gp120 (GWGR). Essa freqüência não é observada em outros subtipos e nem mesmo no subtipo B em outras epidemias de aids no mundo. Para explorar mais as características das variantes GWGR e GPGR, como já dito, foi realizada uma análise detalhada incluindo inferências filogenéticas, análise de diversidade genética e de pressão seletiva em seqüências dos genes *env*, *pol* e *gag*.

Após uma análise preliminar para exclusão de amostras recombinantes e subtipos não B que poderiam influenciar as análises, inicialmente foi realizada uma análise filogenética com todas as amostras selecionadas para se determinar o padrão de agrupamento. A árvore do gene *env* (**Figura 18**) mostrou que as seqüências que contém W no *loop* V3 não formam nenhum agrupamento monofilético dentro do subtipo B, as seqüências independentemente de possuírem o aminoácido W ou P distribuem-se aleatoriamente na árvore de *env*.

A árvore construída para o gene *pol* com as variantes GWGR e GPGR (**Figura 19**) revelou que as amostras se agrupam no subtipo B e ambas as variantes GWGR e GPGR se misturam. No entanto, essa árvore demonstrou a presença de dois grupos de amostras. É provável que essa separação deva-se ao fato de que amostras sob tratamento anti-retroviral foram utilizadas na análise. Isso poderia estar relacionado à maior pressão seletiva em alguns códons, o que poderia influenciar a distribuição das amostras na árvore. Porém, pode-se visualizar tanto amostras GWGR e GPGR misturadas em ambos os grupos de *pol*.

A inferência filogenética da região *gag* (**Figura 20**) também demonstrou que as amostras GWGR e GPGR não diferenciam-se nessa região.

Todas as inferências realizadas até esse momento já sugeriam que amostras GWGR e GPGR não constituem duas variantes distintas do subtipo B brasileiro.

A seguir foram realizadas análises para se calcular a diversidade genética e a pressão seletiva (razão entre dN/dS) usando-se um método de comparação par-a-par. A diversidade genética (**Tabela 15**), além de representar um grande desafio para a elaboração de uma vacina efetiva contra a aids (Esparza *et al.*, 1996) está relacionada à capacidade de variantes mutantes escaparem do sistema imune do hospedeiro, potencialmente propiciarem resultados falso negativos em testes sorológicos convencionais para o seu diagnóstico (McKnight & Clapham,

1995), podendo também estar associada ao surgimento de resistência aos quimioterápicos anti-retrovirais (Najera *et al.*, 1995). Além disso, a evolução viral poderia levar a mudanças na virulência das variantes afetando o ritmo, a transmissão e a progressão da doença (Weiss, 1993). Especula-se que a complexidade genética elevada do HIV-1 pode, em certo momento, dismantelar o controle da replicação viral. (Nowak *et al.*, 1991).

A diversidade genética do HIV-1 pode ser entendida como resultante da pressão seletiva positiva exercida pela resposta imune do hospedeiro, a qual pode mudar de intensidade segundo a região genômica viral, sendo as regiões do *env* as mais expostas a essa pressão do hospedeiro (Brown & Menaghan, 1998). A intensidade dessa pressão seletiva pode ser determinada pela razão entre o número de mutações não sinônimas e sinônimas em cada gene (Kimura *et al.*, 1980; Nei & Gojobori, 1986). Em caso de pressão seletiva neutra a relação dN/dS é 1, sendo que quando há um predomínio de alterações não sinônimas a pressão é positiva ou maior do que 1 e quando há um predomínio de mutações sinônimas a pressão é negativa ou menor do que 1.

De acordo com as análises de pressão seletiva baseadas em mensuração par a par (Nei & Gojobori) encontramos uma pressão seletiva com valores de  $\omega$  próximos de 1,5 na região do *env* tanto para amostras GWGR como GPGR sugerindo mais uma vez uma semelhança nas pressões sofridas nesses genes de ambas as variantes em questão. Na região *gag* foram encontrados valores de  $\omega$  também próximos de 1,5 em amostras GWGR e GPGR também sugerindo que não existem diferenças significativas nas pressões seletivas dessas amostras. Na região *pol* as amostras GWGR apresentaram um valor de  $\omega$  também similar a 1,5, porém as amostras GPGR mostraram valores de  $\omega$  acima de 2. Os valores de  $\omega$  na região *pol* muito provavelmente refletem a presença de códons de resistência medicamentosa em pacientes sob tratamento anti-retroviral. Dessa forma, os valores de  $\omega$  em *pol* apresentam-se mais elevados do que se esperaria encontrar em genes não fortemente imunogênicos como o *pol*. As aferições de pressão seletiva feitas pelo método de mensuração par a par podem gerar imprecisões porque em uma seqüência de aminoácidos existem códons sob pressão seletiva diferenciada, o que não transparece no método par a par porque ele apenas apresenta uma média de todos os valores de  $\omega$  do alinhamento.

Para se fazer os cálculos da diversidade genética e a média da razão entre dN/dS, as amostras foram separadas em diferentes grupos, de acordo com gene e o motivo no *loop* V3 (GPGR ou GWGR). Os resultados dessa análise mostraram os valores da razão entre dN/dS entre os genes são distintos (**Tabela 14**). Os valores de dN/dS se mostraram ligeiramente acima de 1, como discutido acima, indicando a possibilidade de seleção positiva nessas seqüências. Como falado anteriormente, especula-se que em *pol* a pressão esteja associada aos medicamentos anti-retrovirais, pois as amostras do estudo pertencem a indivíduos que tomaram anti-retroviral. Não



fizemos essa exclusão para tentar relatar aquilo que circula na população sob terapia anti-retroviral.

Como mencionado anteriormente, as medidas de diversidade foram baseadas em comparação par a par das distâncias entre as seqüências usando-se a matriz de substituição Kimura 2 Parâmetros, que dá um peso duas vezes maior para as transversões em relação às transições. Para tal análise também separamos as seqüências de acordo com as regiões gênicas e com as variantes GWGR e GPGR. Nas amostras GPGR foi encontrada uma diversidade de 0,36 na região do *env* e de 0,01 e 0,02 nas regiões *gag* e *pol*, respectivamente. Nas amostras GWGR encontramos uma diversidade de 0,18 no *env* e 0,22 e 0,26 nas regiões *gag* e *pol*, respectivamente. Embora nas amostras GPGR os resultados encontrados indicam um maior diversidade em *env*, que é justamente a região mais exposta à pressões seletivas do hospedeiro, isso não foi encontrado em amostras GWGR. Uma possível explicação para esses resultados é, como já dito, termos usado amostras de pacientes sob terapia anti-retroviral.

Para uma aferição mais precisa da pressão seletiva sugerida pelo método par-a-par foram realizadas análises usando um método de verossimilhança que estima os valores de dN/dS em cada códon individualmente presente no alinhamento de seqüências. Os resultados obtidos com essa análise indicaram códons sob seleção positiva ao longo de toda a seqüência da região *env* do HIV-1 em ambas as variantes. Cabe destacar que códons presentes na coroa do *env*, resíduos 306 e 320 apresentaram pressão seletiva em ambas variantes GWGR e GPGR (**Tabela 16**). Isso pode ser influência do próprio sistema imune e talvez essa pressão possa influenciar na escolha de algum tropismo preferido pelo vírus em se ligar a co-receptores R5 ou X4 independente desses vírus possuírem coroa GWGR ou GPGR. Para ambas as variantes foi detectada pressão seletiva nos mesmo códons (**Tabela 16**) indicando uma ausência de pressão seletiva diferenciada entre as variantes. O mesmo modelo evolutivo, o M2, foi escolhido para ambos conjuntos de variantes, sendo os valores de  $\omega$  bem parecidos, em torno de 1,2 para amostras GWGR e 1,3 para amostras GPGR. Foram detectados 7 códons sob pressão seletiva em amostras GWGR e GPGR. Em *gag* foi detectada uma menor quantidade de códons sob pressão positiva em ambas as variantes. Tanto em *gag* de amostras GWGR como GPGR os códons estão sob pressão positiva e em ambas as variantes os valores de  $\omega$  em *gag* foram de 1,4. O modelo selecionado foi o M2. Mais uma vez sendo demonstrada a total consonância de pressão seletiva entre GWGR e GPGR. Alguns estudos sugerem que a presença de resposta citotóxica direcionada aos diferentes epítomos do gene *gag* esteja associada a uma progressão lenta para aids (Riviere *et al.*, 2005) (**Tabela 16**). Na região *pol* foram identificados muitos códons sob seleção positiva. Nas amostras GWGR foram identificados 12 códons sob pressão seletiva sendo a grande maioria códons de resistência aos anti-retrovirais.

Nas amostras GPGR foram identificados 9 códons sob pressão seletiva também correspondendo a códons de resistência anti-retroviral. Os valores de  $\omega$  foram de 1,40 para amostras GWGR e 2,26 para amostras GPGR, sendo o modelo evolutivo escolhido, M2, o mesmo para ambas as variantes. Os valores elevados de  $\omega$  em *pol* deve-se ao fato de que códons de resistência foram analisados. Nessa região fica também demonstrado que amostras GWGR e GPGR não se encontram sob pressões seletivas diferenciadas. Já que a maioria dos códons sob seleção positiva são códons de resistência, o aumento de  $\omega$  em amostras GPGR quando comparado a amostras GWGR, deve-se muito mais a presença e intensidade do tratamento anti-retroviral do que a diferenças genéticas entre as variantes que suscitem pressões diferenciadas (**Tabela 16**).

Estudos realizados *in vitro*, evidenciaram que a resposta de CTL está direcionada principalmente contra poucos epítomos dos genes *env* e *pol*, sugerindo que a progressão tende a ser rápida (Hay *et al.*, 1999). Esses estudos em *pol* e *gag* incentivaram-nos a expandir este estudo para além do gene *env*, incluindo os genes *pol* e *gag*, pois explorar grandes trechos de seqüências virais de outras regiões para verificar se a seqüência GWGR é apenas uma assinatura da alça V3 ou uma real variante, possibilitam a obtenção de dados mais precisos e fidedignos que permitam um melhor entendimento da epidemia HIV/aids no Brasil.

Finalmente, para realmente definir se amostras GWGR e GPGR são variantes distintas ou amostras que contêm apenas uma assinatura diferenciada em *env*, foi composta uma seqüência concatêmica e comparada a distribuição de ambas as amostras em um agrupamento filogenético. Para tal, foram unidas as regiões *gag*, *pol* e *env* de cada amostra GWGR e GPGR, compondo dessa forma um “mini-genoma” de quase 3.000 nucleotídeos, o que dá uma maior robustez para a distribuição encontrada. Ao se fazer essa análise ficou claro que GWGR não são variantes distintas do subtipo B brasileiro, mas apenas seqüências com assinaturas diferentes. A distribuição encontrada demonstrou que não houve separação das amostras GWGR e GPGR e sim uma total mistura entre elas (**Figura 21**).

Essas análises mostraram que os vírus que contêm W ou P no *loop* V3 não estão evoluindo diferencialmente na epidemia de aids no Brasil e não são variantes distintas. Possivelmente, os vírus que contêm a variante GWGR entraram cedo na epidemia (efeito fundador) (Lowagie *et al.*, 1993), sofrendo disseminação no Brasil. Muito provavelmente as diferenças entre GWGR e GPGR não acarretaram mudanças significativas na adaptação desses vírus na população brasileira, o que justifica suas elevadas prevalências até o momento, próximo de 50%.

As análises relevaram que não existe uma distinção entre os variantes GW e GP e que ambos estão sujeitos às mesmas restrições evolutivas na epidemia HIV/aids do Brasil. Assim, os

dados obtidos mostram que o principal fator seletivo no nível populacional é a pressão imune e que a presença ou ausência de motivos W ou P não influi consideravelmente nas estimativas.

## 7.0 – CONCLUSÕES

- Amostras GWGR e GPGR não se diferenciam em árvores filogenéticas de três grandes regiões do genoma: *env*, *pol* e *gag*;
- A diversidade genética de amostras GPGR é maior que a diversidade genética de GWGR no *env*. A diversidade genética de amostras GPGR é menor que a diversidade genética de GWGR em *gag*. A diversidade genética de amostras GPGR é menor que a diversidade genética de GWGR no *pol*;
- Amostras GWGR e GPGR não possuem pressão seletiva diferenciada de acordo com a análise par a par. Amostras GWGR e GPGR não se diferenciam na medida de pressão seletiva códon a códon;
- Amostras GWGR e GPGR não se diferenciam após a montagem de concatêmeros em árvores filogenéticas;
- Os vírus que contém W ou P no *loop* V3 não estão evoluindo diferencialmente na epidemia de aids no Brasil e não são variantes distintas. Possivelmente, os vírus que contêm a variante GWGR entraram cedo na epidemia (efeito fundador) sofrendo disseminação no Brasil;
- Muito provavelmente as diferenças entre GWGR e GPGR não acarretaram mudanças significativas na adaptação desses vírus na população brasileira, o que justifica suas elevadas prevalências até o momento.

## 8.0 – REFERÊNCIAS

AMMANN, A.J. *et al.* Acquired immunodeficiency in an infant: possible transmission by means of blood products. **Lancet**, 1: 956-958, 1983.

ARENS M. Methods for subtyping and molecular comparison of human viral genomes. **Clinical Microbiology Reviews**, 12: 612-626, 1999.

BARRÉ-SINOUSSE, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). **Science**, 220 868-71, 1983.

BONGERTZ V.; BOU-HABIB D. C.; BRÍGIDO L. F.; CASEIRO M.; CHEQUER P. J.; COUTO-FERNANDEZ J. C.; FERREIRA P. C.; GALVAO-CASTRO B.; GRECO D.; GUIMARAES M. L.; LINHARES DE CARVALHO MI, MORGADO M. G.; OLIVEIRA C. A.; OSMANOV S.; RAMOS C. A.; ROSSINI M.; SABINO E.; TANURI A.; UEDA M. HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian network for HIV isolation and characterization. **J. Acquir. Immune Defic. Syndr.**, 23: 184-93, 2000.

BOUR, S. STREBEL, K. The HIV vpr protein: a multifunctional enhancer of viral particle release. **Microbes Infect.** 5: 1029-39, 2003.

BRINDEIRO, R. M.; DIAZ, R. S.; SABINO, E. C.; MORGADO, M. G.; PIRES, I. L.; BRÍGIDO, L.; DANTAS, M. C.; BARRÉIRA, D.; TEIXEIRA, P. R.; TANURI, A. Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. **Aids** 17 (7), 1063-9, 2003.

BRITO, A. M.; CASTILHO, E. A.; SZWARCOWALD, C. L.; AIDS and HIV infection in Brazil: a multifaceted epidemic. **Revista da Sociedade Brasileira de Medicina Tropical.** 34: 207-217, 2000.

BROWN A.L.; MONAGHAN P. Evolution of the structural proteins of human immunodeficiency virus: selective constraints on nucleotide substitution. **Aids Res Hum Retroviruses**, 4: 399-407, 1998.

CASSEB, J.; HONG, M. A.; GONSALEZ, C.; BRÍGIDO, L. F.; DUARTE, A. J.; MICHAEL-HENDRY, R. Two variants of HIV-1 B serotype and transmitted heterosexually in São Paulo, Brazil. **Braz J Med Biol Res**, 31: 1243-1246, 1998.

CHESEBRO, B. *et al.* Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T leukemia cells, and macrophages. **J Virol**, v.65, p. 5782-5789, 1991.

CLAPHAM, P. R.; WEISS, R. A. Immunodeficiency viruses. Spoilt for choice of co-receptors. **Nature**, 388 (6639), 230-1, 1997.

COFFIN, J. M. Structure and classification of retrovirus. In: LEVY, J. A. The Retroviridae. New York: **Plenum Press**, 1992.

COFFIN, J. M., KNIPE, B. N., HOWLEY, D. M. *et al.* The viruses and their replication. Retroviridae: Eds. Fields., Lippincott-Raven publisher, In: **Virology** 1767-1846, 1996.

COLSON, E. R. (AZT) for treatment fo patients infected with human immunodeficiency virus type 1. An evaluation of effectiveness in clinical practice. **Arch Intern Med.** 151: 709-13, 1991.

COUTO-FERNANDEZ, J. C. *et al.* Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: assessing subtype and drug-resistence associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy. **Mem Inst Oswaldo Cruz.** 100: 73-8, 2005.

CULLEN, B. R.; GREENE, W. C. Functions of the auxiliary gene products of the human immunodeficiency virus type 1. **Virology**, 178 (1), 1-5, 1990.

D'SOUZA, M. P.; HARDEN, V. A. Chemokines and HIV-1 second receptors. Confluence of two fields generates optimism in aids research. **Nat Med**, 2: 1293-1300, 1996.

DA COSTA, S. M.; SCHECHTER, M.; SHINDO, N.; VICENTE, A. C.; OLIVEIRA, E. F.; PINTO, M. E.; TANURI, A. Sequence and phylogenetic analysis of glycoprotein 120 of an HIV type 1 variant (GWGR) prevalent in Brazil. **Aids Res Hum Retroviruses**, Sep;11 (9):1143-5, 1995.

DAS, S. R.; JAMEEL, S. Biology of the HIV nef protein. **Indian J Med Res.** 121: 315-32, 2005.

DE CLERCQ, E. New perspectives for the treatment of HIV infections. **Verh K Acad Geneeskd Belg.** 60: 13-45, 1998.

DENG, H.K. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. **Nature** 381: 661-6, 1996.

DE JONG, J. J.; GOUDSMIT, J.; KEULEN, W.; KLAVER, B.; KRONE, W.; TERSMETTE, M.; DE RONDE, A. Human Immunodeficiency Virus Type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. **Journal of Virology**, 66: 757-765, 1992.

DE SÁ FILHO, D. J. *et al.* Identification of two HIV type 1 circulating recombinant forms in Brazil. **Aids Res Hum Retrov.** 22: 1-13, 2006.

DIAZ, R. S. *et al.* Dual human immunodeficiency virus type 1 infection and recombination inadually exposed transfusion recipient. The transfusion safety study group. **J. Virol**, 69: 3273-81, 1995.

DIAZ R.S.; ZHANG M.P.; BUSCH J.W.; MOSLEY J.W.; MAYER A. Divergence of HIV-1 quasispecies in an epidemiologic cluster. **Aids**, 11: 415-22, 1997.

DOMS, R. W. Protein conformational changes in virus-cell fusion. **Methods Enzymol**, 221: 61-72, 1993.

ESPARZA J.; HEYWARD W.L.; OSMANOV S. HIV vaccine development from basic research to human trials. **Aids**, 10 (suppl): 5 123-32, 1996.

FOUCHIER, R. A. *et al.* Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp 120 molecule. **Journal of Virology**, 66: 3183-3187, 1992.

FOUCHIER, R. A. M.; BROUWER, M.; BROERSEN, S. M.; SCHUITEMAKER, H. Simple determination of Human Immunodeficiency Virus Type 1 syncytium-inducing V3 genotype by PCR. **Journal of Clinical Microbiology**, 33: 906-911, 1995.

GOFF, S.P., Death by deamination: a novel host restriction system for HIV-1. **Cell** 114:281-3, 2003.

GROENINK, M. *et al.* Relation of phenotype evolution of HIV-1 to envelope V2 configuration. **Science**, 260: 1513-1516, 1993.

HAY, C. M. *et al.* Lack of viral escape and defective *in vivo* activation of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes in rapidly progressive infection. **J Virol**. 73: 5509-5519, 1999.

HENDRY, R. M.; HANSON, C. V.; BONGERTZ, V.; MORGADO, M.; DUARTE, A.; CASSEB, J.; BRÍGIDO, L.; SABINO, E. C.; DIAZ, R.; GALVÃO-CASTRO, B. Immunoreactivity of Brazilian HIV isolates with different V3 motifs. **Memórias do Instituto Oswaldo Cruz**, 91: 347-348, 1996.

HO, D. D. HIV-1 dynamics *in vivo*. **J Biol Regul Homeost Agents**, v 9, 76-77, 1995.

HOFFMAN, N.G.; Seillier-Moiseiwitsch, F.; Ahn, J.; Walker, J.M.; Swanstrom, R., **J Virol**. 76(8):3852-64, 2002.

JAFEE, H. W.; BREGMAN, D. J.; SELIK, R. M. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. **J Infect Dis**, 148: 399-345, 1983.

JANINI, M. *et al.* Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4 (+) T cells. **J Virol**, 75 (17), 7973-86, 2001.

JETZT, A. E.; YU, H.; KLARMANN, G. J.; RON, Y.; PRESTON, B. D.; DOUGHERTY, J. P. High rate of recombination throughout the human immunodeficiency virus type 1 genome. **J Virol** 74(3), 1234-40, 2000.

KIMURA M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotides sequences. **J Mol Evol.**, 16:111-120, 1980.

KLESTER, H. W. *et al.* Importance of the nef gene for maintenance of high virus loads and for development of aids. **Cell**, 65 (4), 651-62, 1991.

LEVY, J. A. Infection by human immunodeficiency virus--CD4 is not enough. **N Engl J Med**, 335 (20), 1528-30, 1996.

LEVY, J. A. Pathogenesis of human immunodeficiency virus infection. **Microbiol Rev**, 57 (1), 183-289, 1993.

LOWAGIE, J.; McCUTCHAN, F. E. & MASCOLA J. Genetic subtypes of HIV-1, **aids Res. Hum. Retrov.** 9, pp. S147-S150, 1993.

MANSKY L.M.; TEMIN, H.M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than what predicted from the fidelity of purified reverse transcriptase. **J Virol**, v. 69, 5087-5094, 1995.

McCUSKER J. *et al.* Readmissions to drug abuse treatment and HIV risk behavior. **Am J Drug Alcohol Abuse**, 24: 523-540, 1998.

McCUTCHAN, F. E. Understanding the genetic diversity of HIV-1. **Aids**, 14: 31-44, 2000.

McKEATING J.A.; GOW J.; GOUDSMIT J.; PEARL L.H.; MULDER C.; WEISS R.A. Characterization of HIV-1 neutralization escape mutants. **Aids**, 3: 777-84, 1989.

McKNIGHT A.; CLAPHAM P.R. Immune escape and tropism of HIV. **Trends microbial**, 3: 356-61, 1995.

MILICH L.; MARGOLIN, B.; SWANSTROM R. V3 loop of the human immunodeficiency virus type 1 env protein: interpreting sequence variability. **J Virol**, 67, 5623-34, 1993.

MORGADO M.G.; SABINO E.C.; SHAPAER E.G, *et al.* V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from the North American/European prototype and detection of subtype F. **aids Res Hum Retroviruses** 10: 567-569, 1994.

MORGADO, M.G.; GUIMARÃES, M. L.; GRIPP, C. B. G.; COSTA, C.I.; JUNIOR, I.N.; VELOSO, V. G.; CARVALHO, M. I. L.; BRANCO, L. R. C.; BASTOS, F.I.; KUKEN, C.; CASTILHO, E. A.; CASTRO, B. G.; BONGERTZ, V.; CHAGAS, E. Molecular epidemiology of HIV-1 in Brazil: High prevalence of HIV-1 subtype B and identification of HIV-1 subtype D infection in the Rio de Janeiro, Brazil. **Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology**, 18: 488-494, 1998.

MUSE, S. V.; GAUT, B. S. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. **Mol Biol. Evol.** 11: 715-724, 1994.



NAJERA I. *et al.* Pol gene quasispecies of immunodeficiency virus mutations associated with drug resistance in virus from patients undergoing no drug therapy. **J. Virol**, 69: 23-31, 1995.

NEGRONI, M.; BUC, H. Mechanisms of retroviral recombination. **Annu Rev Genet** 35, 275-302, 2001.

NEI, M; GOJOBORI, T. Simple Methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. **Mol Biol Evol**, 3: 418-426, 1986.

NIELSEN, R.; YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**, v. 148, p. 929-936, 1998.

NOWAK, M.A. *et al.* Antigenic diversity thresholds and the development of aids. **Science**, 254: 963, 1991.

OLESKE, J. *et al.* Immune deficiency syndrome in children. **Jama**, 249: 2345-2349, 1983.

POSADA, D.; CRANDALL, K. A. MODELTEST: testing the model of DNA substitution. **Bioinformatics**, 14(9): 817-8, 1998.

POTTS, K. E.; KALISH, M. L.; LOTT, T.; ORLOFF, G.; LUO, C. C.; BERNARD, M. A.; ALVES, C. B., BADARO, R.; SULEIMAN, J.; FERREIRA, O. Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. **Aids**, 7: 1191-1197, 1993.

PRESTON, B. D.; POIESZ, B. J.; LOEB, L. A. Fidelity of HIV-1 reverse transcriptase. **Science**, 242 (4882), 1168-71, 1988.

RIVIERE, Y. *et al.* Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. **Aids Res Hum Retrov.** 11: 903-907, 1997.

ROETH, J. F.; COLLINS, K. L. Human immunodeficiency virus type 1 Nef: adapting to intracellular trafficking pathways. **Microbiol Mol. Biol. Rev.** 70: 548-63, 2006.

RAMBAUT, A. *et al.* The causes and consequences of HIV evolution. **Nat Genet**, v. 5, p. 52-61, 2000.

ROBERTSON, D. L.; ANDERSON, J. P.; BRADAC, J. A.; CARR, J. K.; FOLEY, B.; FUNKHOUSER, R. K.; GAO, F.; HAHN, B. H.; KALISH, M. L.; KUIKEN, C.; LEARN, G. H.; LEITNER, T.; McCUTCHAN, F.; OSMANOV, S.; PEETERS, M.; PIENIAZEK, D.; SALMINEN, M.; SHARP, P. M.; WOLINSKY, S.; KORBER, B. HIV-1 nomenclature proposal. **Science**, 288 (5463), 55-6, 2000.

ROBERTSON D. L. *et al.* Recombination in HIV-1. **Nature**, 374 (6518), 124-6, 1995.

ROSSINI, M. A.; DIAZ, R. S.; CASEIRO, M.; TURCATO, G.; ACCETTURI, C. A., SABINO, E. C. HIV-1 subtypes among intravenous drug users from two neighboring cities in Sao Paulo State, Brazil. **Braz J Med Biol Res**, 34 (1), 45-7, 2001.

SÁ FILHO, D. J.; SANABANI, S. S. M.; DIAZ, R. S.; MUNERATO, P.; BRUNSTEIN, A.; FUSUMA, E. E.; SABINO, E. C.; JANINI, L. M. R. Analysis of full-length human immunodeficiency virus type 1 genome reveals a variable spectrum of subtypes B and F recombinants in São Paulo, Brazil. **Aids Research and Human Retrovirus**, 21: 145-151, 2005.

SABINO, E. C.; DIAZ, R. S.; BRÍGIDO, L. F.; LEARN, G. H.; MULLINS, J. I.; REINGOLD, A. L.; DUARTE, A. J.; MAYER, A.; BUSCH, M. P. Distribution of HIV-1 subtypes seen in an aids clinic in Sao Paulo City, Brazil. **Aids** 10 (13), 1579-84, 1996.

SABINO, E. C.; SHPAER, E. G.; MORGADO, M. G.; KORBER, B. T.; DIAZ, R. S.; BONGERTZ, V.; CAVALCANTE, S.; GALVÃO-CASTRO, B.; MULLINS, J. I.; MAYER, A. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. **J Virol** 68 (10), 6340-6, 1994.

SALMINEN, M. O.; CARR, J. K.; BURKE, D. S. and McCUTCHAN, F. E. Identification of break-points in intergenotypic recombinants of HIV type 1 by bootscanning. **Aids Res Hum Retroviruses**. 11(11): 1423-25, 1995.

SANABANI, S. *et al.* Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in Sao Paulo, Brazil. **Infect Genet Evol.** 6: 368-77, 2006.

SANTORO-LOPES, G.; HARRISON, L. H.; TAVARES, M. D.; XEXEO, A.; SANTOS, A. C.; SCHECHETER, M. V3-binding serotypes and HIV disease progression in Brazil: Is B different from B<sup>2</sup>. **Aids Res Hum Retrov**, 16: 953-8, 2000.

SHANKARAPPA, R. *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. **J. Virol.** 73 (12): 10489-502, 1999.

SHERIDAN, P. L.; SHELINE, C.; MILOCO, S. H.; JONES, K. A. Tat and the HIV promoter: a model for RNA-mediated regulation of transcription. **Semin Virol.** 4: 69-80, 1991.

SIMON F.; MAUCLERE P.; ROQUES P.; LOUSSERT-AJAKA I.; MULLER TRUTWIN M.C.; SARAGOSTI S, *et.al.* Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. **Nat. Med**, 4: 1032-7, 1998.

SOARES, M. A.; OLIVEIRA, T.; BRINDEIRO, R. M.; DIAZ, R. S.; SABINO, E. C.; BRÍGIDO, L.; PIRES, I. L.; MORGADO, M. G.; DANTAS, M. C.; BARRÉIRA, D.; TEIXEIRA, P. R.; CASSOL, S.; TANURI, A. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. **Aids** 17 (1), 11-21, 2003.

STREBEL, K. *et al.* The HIV 'A' (sor) gene product is essential for virus infectivity. **Nature**, 328(6132), 728-30, 1987.

SWOFFORD, D. L. *et al.* Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. **Syst Biol.** 50:525-39, 2002.

TANURI, A.; SWANSON, P.; DEVARE, S.; BERRO, O. J.; SAVEDRA, A.; COSTA, L. J.; TELLES, J. G.; BRINDEIRO, R.; SCHABLE, C.; PIENIAZEK, D.; RAYFIELD, M. HIV-1 subtypes among blood donors from Rio de Janeiro, Brazil. **J Acquir Immune Defic Syndr Hum Retrovirol** 20 (1), 60-6, 1999.

THOMPSON, J. D. *et al.* The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res**, 25: 4876-82, 1997.

THOMSON M. M.; PEREZ-ALVAREZ, L.; NAJERA, R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. **Lancet Infect Dis.** 2:461-471, 2002.

THOMSON, M. M. *et al.* Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. **Lancet Infect Dis** 2 (8), 461-71, 2002b.

TUNGATURTHI, P.K. *et al.* Role of HIV vpr in aids pathogenesis: relevance and implications of intravirion, intracellular and free vpr. **Biomedicine & Pharmacotherapy.** 57:20-24, 2003.

VARTANIAN, J. P. *et al.* Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. **J Virol**, 65 (4), 1779-88, 1991.

VICENTE, A. C. *et al.* The HIV epidemic in the Amazon Basin is driven by prototypic and recombinant HIV-1 subtypes B and F. **J Acquir Immune Defic Syndr**, 23: 327-331, 2000.

WEISS R.A. How does HIV cause aids? **Science**, 260: 1283-9, 1993.

WELLENSIEK, B. P.; SUNDARAVARADAN, V.; RAMAKRISHNAN, R.; AHMAD, N. Molecular characterization of the HIV-1 gag nucleocapsid gene associated with vertical transmission. **Retrovirology.** 3: 21, 2006.

YANG, Z. *et al.* Statistical properties of the maximum likelihood method of the phylogenetic estimation and comparison with distance matrix methods. **Syst Biol.** 43: 329-342, 1994.

YANG, Z. *et al.* Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. **Mol Biol Evol:**15:568-573,1997.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)