

Universidade Federal da Bahia  
Instituto de Ciências da Saúde



**PMBqBM**  
**Programa Multicêntrico de**  
**Pós-Graduação em Bioquímica e Biologia Molecular**

**Juliana Sacramento Mota de Souza**

**CARACTERIZAÇÃO MOLECULAR E GEOGRÁFICA DE ISOLADOS  
RECOMBINANTES BF DO VÍRUS DA IMUNODEFICIÊNCIA  
HUMANA DO TIPO 1 (HIV-1)**

Salvador  
2017



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE CIÊNCIAS DA SAÚDE  
DEPARTAMENTO DE BIOQUÍMICA E BIOFÍSICA  
PROGRAMA DE PÓS-GRADUAÇÃO MULTICÊNTRICO EM  
BIOQUÍMICA E BIOLOGIA MOLECULAR (PMBqBM/UFBA-SBBq)**



**JULIANA SACRAMENTO MOTA DE SOUZA**

**CARACTERIZAÇÃO MOLECULAR E GEOGRÁFICA DE ISOLADOS  
RECOMBINANTES BF DO VÍRUS DA IMUNODEFICIÊNCIA  
HUMANA DO TIPO 1 (HIV-1)**

**SALVADOR**

**2017**

**JULIANA SACRAMENTO MOTA DE SOUZA**

**CARACTERIZAÇÃO MOLECULAR E GEOGRÁFICA DE ISOLADOS  
RECOMBINANTES BF DO VÍRUS DA IMUNODEFICIÊNCIA  
HUMANA DO TIPO 1 (HIV-1)**

Dissertação apresentada ao Programa Multicêntrico de Pós-Graduação em Bioquímica e Biologia Molecular – UFBA/SBBq, como requisito parcial para a obtenção do título de Mestre em Bioquímica e Biologia Molecular pela Universidade Federal da Bahia.

Orientadora: Dr.<sup>a</sup> Joana Paixão Monteiro Cunha

**SALVADOR**

**2017**

Modelo de ficha catalográfica fornecido pelo Sistema Universitário de Bibliotecas da UFBA para ser confeccionada pelo autor

SO729 SOUZA, Juliana Sacramento Mota de  
CARACTERIZAÇÃO MOLECULAR E GEOGRÁFICA DE ISOLADOS  
RECOMBINANTES BF DO VÍRUS DA IMUNODEFICIÊNCIA HUMANA DO TIPO  
1 (HIV-1) / Juliana Sacramento Mota de SOUZA. -- Salvador,  
2017.  
128 f. : il

Orientadora: Joana Paixão Monteiro Cunha.  
Dissertação (Mestrado - Multicêntrico em Bioquímica e  
Biologia Molecular) -- Universidade Federal da Bahia,  
Universidade Federal da Bahia, 2017.

1. HIV-1. 2. Recombinantes. 3. BF. 4. Caracterização  
molecular. 5. Caracterização geográfica. I. Cunha, Joana Paixão  
Monteiro. II. Título.

**JULIANA SACRAMENTO MOTA DE SOUZA**

**CARACTERIZAÇÃO MOLECULAR E GEOGRÁFICA DE ISOLADOS  
RECOMBINANTES BF DO VÍRUS DA IMUNODEFICIÊNCIA HUMANA  
DO TIPO 1 (HIV-1)**

Dissertação apresentada ao Programa Multicêntrico de Pós-Graduação em Bioquímica e Biologia Molecular (PMBqBM), Instituto de Ciências da Saúde, Universidade Federal da Bahia, como requisito para a obtenção do título de Mestre em Bioquímica e Biologia Molecular.

Salvador, 09 de março de 2017.

**BANCA EXAMINADORA**



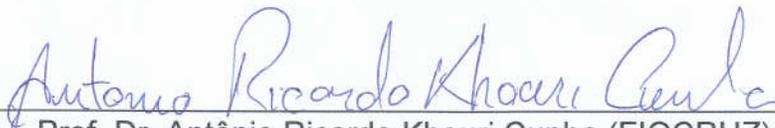
---

Profa. Dra. Joana Paixão Monteiro Cunha (Orientadora/UFBA)  
(Presidente)



---

Profa. Dra. Glória Regina Franco (UFMG)



---

Prof. Dr. Antônio Ricardo Khouri Cunha (FIOCRUZ)

## **AGRADECIMENTOS**

À Prof<sup>ª</sup>. Dr<sup>ª</sup>. Joana Paixão Monteiro Cunha pela oportunidade, orientação, confiança, paciência e incentivo.

Aos colegas do Núcleo de Bioinformática, Departamento de Biofunção, da Universidade Federal da Bahia, pela oportunidade de convívio e apoio.

Ao Programa de Pós-Graduação Multicêntrico em Bioquímica e Biologia Molecular, pela qualidade da capacitação ofertada.

Aos colegas de curso, pelo companheirismo nesta jornada e pela agradável convivência.

À todos que direta ou indiretamente fizeram parte da minha formação.

À FAPESB, pelo fomento a meu aperfeiçoamento científico e acadêmico.

À minha família, base na qual pude e posso me fortalecer a cada dia. Em especial a meus pais, que sempre incentivaram minha dedicação aos estudos.

Aos meus amigos, pelo apoio em todos os momentos acadêmicos e pessoais.

À Deus, pela força na qual pude e posso me sustentar a cada dia.

## RESUMO

O Vírus da imunodeficiência humana do tipo 1 (HIV-1) possui uma ampla variabilidade genética. Esta diversidade é representada pelos quatro grupos, pelos nove subgrupos do grupo M e também pelas formas recombinantes destes vírus. Dentre estes, os recombinantes BF tem se destacado pela alta dispersão global concomitante com o aumento em número e diversidade. Atualmente quatorze formas recombinantes circulantes (CRFs) BF e inúmeras formas recombinantes únicas BF (URFs) foram descritas. O objetivo deste trabalho foi realizar a caracterização molecular e geográfica de isolados virais recombinantes BF do HIV-1. Para tal, sequências e informações de genomas completos de recombinantes BF do HIV foram coletadas dos bancos de dados NCBI e Los Alamos. Posteriormente foram realizadas a caracterização destas sequências, análises filogenéticas, de recombinação e da alça V3 da proteína glicoproteína 120 (gp120), além de busca por assinaturas nucleotídicas e predição do coreceptor de entrada viral. Foram encontradas 252 sequências de genoma completo (>7000 pb) de recombinantes BF do HIV oriundas de treze países diferentes, sendo a maioria delas provenientes de pacientes brasileiros (52,8%). Seis sequências previamente caracterizadas como recombinantes BF, foram reclassificadas como subtipo B puro. Algumas sequências formaram agrupamentos monofiléticos com as CRFs já descritas e, portanto, sugere-se a reclassificação das mesmas. Dentre as sequências avaliadas foram encontrados 114 padrões distintos de recombinação. A maioria das sequências apresenta recombinação na região *pol* (81%; 205/252), seguido das regiões *gag* (68%; 172/252) e *env* (54%; 137/252) sendo encontrado na região *tat* a menor taxa de recombinação (7%; 17/252). Enquanto em todos os outros genes acessórios e regulatórios predominou o genótipo B, em *vif*, o subtipo F prevaleceu (38%; 96/252). Além disso, foram identificados dois *hotspots* entre as sequências analisadas: um encontrado em 35 sequências (13,9%) na região entre as posições 5360 – 5390 do gene acessório *vif* e outro em torno da posição 9356 da região *nef* (11% das sequências). Foram identificadas assinaturas nucleotídicas conservadas exclusivas para cada CRF\_BF descrita. As mutações mais frequentes foram encontradas na região da transcriptase reversa foram M41L (NRTIs) e K100N (NNRTIs). De maneira geral, as mutações encontradas nestas sequências conferem resistência a diversos fármacos sendo que as maiores taxas de resistência foram encontradas para os antirretrovirais Atazanavir, Saquinavir, Nevirapina, Zidovudina, Estavudina, Didanosina e Emtricitabina. A predição de uso de coreceptor viral mostrou estes recombinantes utilizam preferencialmente o coreceptor CCR5 (67,5%) sendo que a CRF42 utiliza exclusivamente este coreceptor, enquanto todas as sequências da CRF39 foram classificadas como R5X4/X4. Os tetrapeptídeos mais frequentes encontrados na alça V3 foram GPGR (49,4%), GPGQ (20%), APGR (6,7%), GPGK (5,1%), GWGR (4,3%) e GPGG (2%). Além disso, as sequências com o motivo GPGQ e APGR são preferencialmente classificadas como R5 e contem o subtipo F na região da alça V3. Pode-se observar que há uma grande diversidade dos padrões de recombinação evidenciando que a recombinação entre os subtipos B e F é frequente.

**Palavras-chave:** HIV-1. Recombinantes BF. Caracterização geográfica. Caracterização molecular. CRF.

## ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) has a wide genetic variability. This diversity is represented by the four groups, nine subgroups of group M and by the recombinant forms of these viruses. Among these, the recombinant BF has been distinguished by the high global dispersion concomitant with the increase in number and diversity. Currently fourteen BF Circulating Recombinant Forms (CRFs) and BF Unique Recombinant Forms (URFs) forms have been described. The objective of this work was to perform molecular and geographic characterization of HIV-1 recombinant BF viral isolates. To that end, complete genome sequences and information from recombinant HIV BFs were collected from the NCBI and Los Alamos databases. Subsequently, the characterization of these sequences, phylogenetic analysis, recombination and the V3 loop of the glycoprotein 120 protein (gp120) were carried out, as well as the search for nucleotide signatures and prediction of the viral entry coreceptor. A total of 252 complete genome sequences (> 7000 bp) of recombinant HIV BF from thirteen different countries were found, most of them coming from Brazilian patients (52.8%). Six sequences previously characterized as recombinant BF were reclassified as pure B subtype. Some sequences formed monophyletic clusters with the CRFs already described and, therefore, it is suggested to reclassify them. Among the sequences evaluated, 114 distinct patterns of recombination were found. Most of the sequences present recombination in the *pol* region (81%, 205/252), followed by the *gag* (68%, 172/252) and *env* (54%; 137/252) regions being found in the *tat* region at the lowest recombination rate (7%, 17/252). While in all other accessory and regulatory genes, genotype B predominated, in *vif*, the F subtype was dominant (38%, 96/252). In addition, two hotspots were identified among the sequences analyzed: one found in 35 sequences (13.9%) in the region between positions 5360-5390 of the *vif* accessory gene and another around position 9356 of the *nef* region (11% of sequences). Unique conserved nucleotide signatures were identified for each CRF\_BF described. The most frequent mutations were found in the reverse transcriptase region were M41L (NRTIs) and K100N (NNRTIs). In general, the mutations found in these sequences confer resistance to several drugs, and the highest resistance rates were found for the antiretrovirals Atazanavir, Saquinavir, Nevirapine, Zidovudine, Stavudine, Didanosine, and Emtricitabine. The prediction of the use of viral coreceptor showed that these recombinants preferentially use the CCR5 coreceptor (67.5%) and CRF42 exclusively uses this coreceptor, while all CRF39 sequences were classified as R5X4 / X4. The most frequent tetrapeptides found in the V3 loop were GPGR (49.4%), GPGQ (20%), APGR (6.7%), GPGK (5.1%), GWGR (4.3%) and GPGG (%). In addition, the sequences with the GPGQ motif and APGR are preferably classified as R5 and contain the F subtype in the V3 loop region. It can be observed that there is a great diversity of the recombination patterns evidencing that the recombination between the subtypes B and F is frequent.

**Keywords:** HIV-1. BF recombinants. Geographical characterization. Molecular characterization. CRF.

## LISTA DE FIGURAS

<b>Figura 1</b> – Mapa genômico do HIV-1 da cepa HXB2.	<b>15</b>
<b>Figura 2</b> – Esquema da classificação do HIV-1 em grupos (N, O, P e M), subgrupos do grupo M (A-D, F-H, J-K), subsubgrupos do subgrupo A (A1- 4) e F (F1 - 2) e as URFs e CRFs.	<b>17</b>
<b>Figura 3</b> – Ciclo de replicação do HIV.	<b>20</b>
<b>Figura 4</b> – Caracterização geográfica dos recombinantes BF do HIV-1 circulantes no mundo.	<b>34</b>
<b>Figura 5</b> - Árvore filogenética ML com leiaute circular.	<b>36</b>
<b>Figura 6</b> - Cluster da árvore ML retangular contendo as CRFs 38 e 44.	<b>37</b>
<b>Figura 7</b> - Diferentes agrupamentos da árvore ML retangular contendo as CRF 39, 40, 42 e 47.	<b>37-38</b>
<b>Figura 8</b> - Agrupamento da árvore ML retangular contendo a CRF46.	<b>39</b>
<b>Figura 9</b> - Agrupamento da árvore ML retangular contendo a CRF29 e sequências relacionadas.	<b>39</b>
<b>Figura 10</b> - Agrupamento da árvore ML retangular contendo a CRF28 e sequências relacionadas.	<b>40</b>
<b>Figura 11</b> - Agrupamento da árvore ML retangular contendo a CRF72 e sequências relacionadas.	<b>40</b>
<b>Figura 12</b> - Agrupamento da árvore ML retangular contendo a CRF17 e sequências relacionadas.	<b>41</b>
<b>Figura 13</b> - Agrupamento da árvore ML retangular contendo a CRF70, CRF71 e sequências relacionadas.	<b>42</b>
<b>Figura 14</b> - Agrupamento da árvore ML retangular contendo a CRF12 e sequências relacionadas.	<b>43</b>
<b>Figura 15</b> - Sequências classificadas como BF porém com padrão de recombinação B na árvore ML retangular.	<b>45</b>
<b>Figura 16</b> - Padrão de recombinação oficial da CRF40 e o padrão encontrado individualmente para cada sequência que compõe a CRF40.	<b>47</b>
<b>Figura 17</b> – Análise de assinaturas nucleotídicas.	<b>51</b>

## LISTAS DE GRÁFICOS

<b>Gráfico 1</b> – Classificação das sequências.	<b>35</b>
<b>Gráfico 2</b> – Análise da presença de recombinação em cada região do genoma de HIV.	<b>49</b>
<b>Gráfico 3</b> – Porcentagem de susceptibilidade encontrada para as diferentes classes de ARVs estudados.	<b>52</b>
<b>Gráfico 4</b> – Resistência a duas classes de ARVs. A porcentagem indica a frequência de sequências que apresentaram mutações que conferem algum nível de resistência à duas classes combinadas.	<b>53</b>
<b>Gráfico 5</b> – Níveis de resistência encontrados para os diferentes fármacos inibidores da protease.	<b>55</b>
<b>Gráfico 6</b> – Níveis de resistência encontrados para os diferentes fármacos da classe dos NRTIs.	<b>57</b>
<b>Gráfico 7</b> – Níveis de resistência encontrados para os diferentes fármacos da classe dos NNRTIs.	<b>57</b>
<b>Gráfico 8</b> – Níveis de resistência encontrados para os diferentes fármacos da classe dos INIs.	<b>58</b>
<b>Gráfico 9</b> – Prevalência da predição de vírus R5 ou R5X4/X4 de acordo com a classificação.	<b>60</b>
<b>Gráfico 10</b> – Frequência de cada motivo estrutural do tetrapeptídeo da alça V3 da proteína gp120 encontrado nas sequências estudadas.	<b>61</b>
<b>Gráfico 11</b> – Frequência do uso de coreceptor predito de acordo com o motivo estrutural encontrado na alça V3 da proteína gp120.	<b>62</b>
<b>Gráfico 12</b> – Classificação das sequências com os motivos mais frequentes na alça V3 de acordo como subtipo desta região e o coreceptor predito.	<b>62</b>

## LISTA DE TABELAS

<b>Tabela 1</b> – Informações coletadas sobre as sequências. D: dado disponível.	<b>33</b>
<b>Tabela 2</b> – Mutações encontradas na região da protease que conferem algum nível de resistência ao PIs.	<b>54</b>
<b>Tabela 3</b> – Mutações encontradas na região da transcriptase reversa (RT) que conferem algum nível de resistência aos NRTIs ou NNRTIs.	<b>56</b>
<b>Tabela 4</b> – Mutações encontradas na região da integrase que conferem algum nível de resistência ao INIs.	<b>58</b>

## **LISTA DE APÊNDICES**

<b>APÊNDICE A</b> – Tabela com números de acesso no GeneBank das sequências utilizadas neste estudo.	<b>72</b>
<b>APÊNDICE B</b> – Padrões de recombinação das CRFs_BF descritas até o momento.	<b>73</b>
<b>APÊNDICE C</b> – Padrões de recombinação identificados no jpHMM.	<b>75</b>
<b>APÊNDICE D</b> – Assinaturas nucleotídicas conservadas (AMNC) exclusivas de cada CRF.	<b>126</b>

## LISTA DE SIGLAS E ABREVIATURAS

3TC	Lamivudina
ABC	Abacavir
AIC	Critérios de informação de Akaike
AIDS	Síndrome da Imunodeficiência Adquirida
AMNC	Assinaturas nucleotídicas conservadas
AR	Argentina
ARV	Antirretroviral
ATV	Atazanavir
AZT	Zidovudina
BF	HIV recombinantes entre o subtipo B e F
BO	Bolívia
BR	Brasil
CCR5	Receptor de quimiocina CCR5
CH	Chile
cpx	Complexo
CRF	Forma recombinante circulante
CXCR4	Receptor de quimiocina CXCR4
D4T	Estavudina
DDI	Didanosina
DNA	Ácido desoxirribonucleico
DRV	Darunavir
DTG	Dolutegravir
EFV	Efavirenz
ENV	Envelope
ES	Espanha
ETV	Etravirina
EVG	Elvitegravir
FPR	Taxa de falso positivo
FPV	Fosamprenavir
FTC	Emtricitabina
GAG	Grupo antígeno específico
gp120	glicoproteína 120

GTR+G+I	Modelo de substituição geral reversível no tempo + Distribuição gama + sítios invariáveis
HIV-1	Vírus da imunodeficiência humana do tipo 1
HIV-2	Vírus da imunodeficiência humana do tipo 2
HIVdb	Banco de Dados de Resistência a Drogas do HIV da Universidade de Stanford
HXB2	Cepa de referência do HIV-1
IDV	Indinavir
INI	Inibidores da integrase
IT	Itália
JP	Japão
jpHMM	<i>jumping profile Hidden Markov Model</i>
LA	Banco de dados do Laboratório Nacional de Los Alamos
LPV	Lopinavir
LTR	Sequências terminais repetidas longas - <i>Long terminal repeat</i>
LX	Luxemburgo
MC	Macau
ML	Método da máxima verossimilhança
NCBI	Centro Nacional de Informação Biotecnológica
NGS	Sequenciamento de nova geração - <i>Next Generation Sequencing</i>
NJ	Método de Agrupamento de vizinhos
NNI	Algoritmo de troca entre vizinhos mais próximos
NNRTI	Inibidores não nucleosídicos da transcriptase reversa
NRTI	Inibidores nucleosídicos da transcriptase reversa -
NVP	Nevirapina
OMS	Organização mundial da saúde
pb	Pares de bases
PE	Peru
PI	Inibidores da protease
POL	Polimerase
PR	Paraguai
R5	Vírus que utilizam o coreceptor CCR5 no processo de entrada viral
R5/X4	Vírus que utilizam o coreceptor CCR5 ou CXCR4 no processo de

	entrada viral
R5X4/X4	Vírus duplo trópicos ou X4
RAL	Raltegravir
RNA	Ácido ribonucleico
RPV	Rilpivirina
RRE	Elemento responsivo de <i>rev</i>
RT	Transcriptase reversa
SIV	Vírus da imunodeficiência Símia
SQV	Saquinavir
TCD4	Linfócitos T CD4 <sup>+</sup>
TCD8	Linfócitos T CD8 <sup>+</sup>
TDF	Tenofovir
TPV	Tipranavir
UNAIDS	Programa Conjunto das Nações Unidas sobre HIV/AIDS
UR	Uruguai
URF	Forma recombinante única
V3	Terceira alça variável da proteína gp120
VIF	Fator de Infectividade Viral - <i>Viral infectivity fator</i>
VZ	Venezuela
X4	Vírus que utilizam o coreceptor CXCR4 no processo de entrada viral

## SUMÁRIO

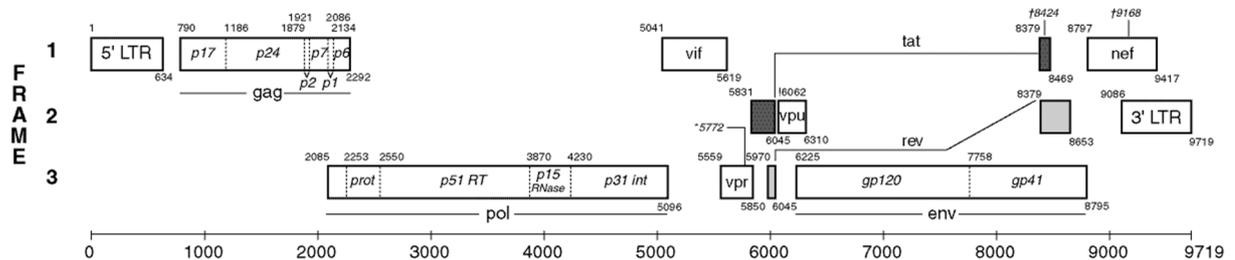
<b>1 INTRODUÇÃO</b>	<b>15</b>
<b>1.1 Classificação</b>	<b>16</b>
<b>1.2 Epidemiologia molecular do HIV-1</b>	<b>17</b>
<b>1.3 Ciclo de replicação</b>	<b>19</b>
<b>1.4 Recombinação e diversidade genética no HIV-1</b>	<b>20</b>
<b>1.5 A alça V3 da proteína gp120</b>	<b>22</b>
<b>1.6 Resistência aos antirretrovirais</b>	<b>22</b>
<b>2 JUSTIFICATIVA</b>	<b>24</b>
<b>3 OBJETIVOS</b>	<b>26</b>
<b>4 MATERIAIS E MÉTODOS</b>	<b>27</b>
<b>4.1 Coleta de dados</b>	<b>27</b>
<b>4.2 Alinhamento e edição</b>	<b>27</b>
<b>4.3 Reconstrução filogenética</b>	<b>28</b>
<b>4.4 Análises de recombinação</b>	<b>29</b>
<b>4.5 Análise das características moleculares</b>	<b>29</b>
<b>4.6 Análise de resistência aos antirretrovirais</b>	<b>30</b>
<b>4.7 Predição de uso do coreceptor</b>	<b>31</b>
<b>4.8 Análise estatística</b>	<b>32</b>
<b>5 RESULTADOS E DISCUSSÃO</b>	<b>33</b>
<b>5.1 Caracterização das sequências</b>	<b>33</b>
<b>5.2 Análise filogenética</b>	<b>35</b>
<b>5.3 Análise de recombinação</b>	<b>43</b>
<b>5.3.1 Reclassificação de sequências previamente classificadas como BF</b>	<b>44</b>

5.3.2	Discrepâncias na identificação dos padrões de recombinação	46
5.3.3	Investigação de pontos preferenciais ( <i>Hotspots</i> ) para recombinação no genoma do HIV-1	47
<b>5.4</b>	<b>Assinatura de nucleotídeos</b>	<b>50</b>
<b>5.5</b>	<b>Análise de resistência aos antirretrovirais</b>	<b>52</b>
5.5.1	Resistência aos PIs	53
5.5.2	Resistência aos NRTIs e NNRTIs	55
5.5.3	Resistência aos INIs	58
<b>5.6</b>	<b>Predição do coreceptor de entrada viral</b>	<b>59</b>
<b>5.7</b>	<b>Reclassificação de sequências</b>	<b>63</b>
<b>5.8</b>	<b>Considerações finais</b>	<b>64</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>66</b>
	<b>REFERÊNCIAS</b>	<b>68</b>

## 1 INTRODUÇÃO

A Síndrome da Imunodeficiência Adquirida (AIDS) é causada pelo HIV e se tornou um problema de saúde pública desde o surgimento da pandemia na década de 80. O HIV é um vírus da família *Retroviridae*, que é constituída por vírus chamados de retrovírus. Os retrovírus possuem um genoma constituído por RNA fita simples senso positivo e replicam o RNA viral através do processo chamado transcrição reversa. Estes vírus são capazes de fazer o fluxo reverso da informação genética: síntese de DNA a partir de moldes de RNA utilizando a enzima viral transcriptase reversa. Outra característica importante destes vírus é a integração do DNA no genoma da célula hospedeira (genoma proviral). Os retrovírus possuem três grupos de genes estruturais: *gag* (antígeno específico), *pol* (polimerase) e *env* (envelope). Na **Figura 1** está representado o mapa genômico do HIV-1, com cerca de 9700 pares de bases (pb) de comprimento, no qual pode-se observar ainda os genes regulatórios *tat* e *rev*; os genes acessórios *vif*, *vpr*, *nef* e *vpu*; e as sequências repetidas nas extremidades 3' e 5' (LTR – *Long terminal repeat*).

**Figura 1** – Mapa genômico do HIV-1 da cepa HXB2.



**Fonte:** Disponível em <http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>

O gene *gag* codifica proteínas do capsídeo viral. Este gene codifica um poliproteína chamada de p55 que será processada pela protease viral nas proteínas p17 (matriz), p24 (capsídeo), p7 (nucleocapsídeo) e p6. O gene *pol* também codifica uma poliproteína que codifica as enzimas virais protease, transcriptase reversa e integrase. Já o gene *env*, que codifica glicoproteínas virais, tem como precursor a proteína gp160 que é processada na glicoproteína transmembrana gp41 e na glicoproteína externa gp120. Esta possui sítios de ligação tanto para o receptor CD4 quanto para coreceptores utilizados no processo da entrada viral. Os genes *tat* e *rev* são genes de proteínas regulatórias da expressão gênica no HIV que interagem com os elementos estruturais genômicos LTR e com o elemento responsivo de *rev* (RRE), respectivamente. Além dos genes estruturais e regulatórios citados, o genoma do HIV

contem genes acessórios que codificam importantes proteínas para o ciclo viral. A proteína *vpr* possui as seguintes funções propostas: direcionar a importação nuclear dos complexos de pré-integração, diminuir o crescimento celular, transativação de genes celulares e indução da diferenciação celular. A proteína viral U, ou *vpu*, exclusiva do HIV-1, é uma proteína integral de membrana que tem duas funções principais: degradar o CD4 no retículo endoplasmático e aumentar a liberação dos vírions. NEF é a proteína mais imunogênica das proteínas acessórias e é essencial para a dispersão viral e para a progressão da doença *in vivo*. Já a proteína acessória *vif* (*Viral infectivity factor* – Fator de Infectividade Viral) que promove a infectividade viral, impede a ação da proteína celular APOBEC-3G, que tem a capacidade de tornar o HIV-1 não infeccioso através da indução de mutações G para A durante a síntese do DNA viral (<https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>).

O HIV pertence ao gênero *Lentiviridae* que é caracterizado por causar infecções crônicas e latentes por um longo período de tempo. A hipótese mais aceita sobre a origem do HIV é que o mesmo teria sido transmitido ao homem através de fluidos de primatas não-humanos infectados com outro vírus deste mesmo gênero, o Vírus da Imunodeficiência Símia (SIV). Dados divulgados pela UNAIDS em 2016 mostram que há uma estimativa de 36,7 milhões de pessoas infectadas como HIV em todo o mundo e que cerca de 35 milhões de pessoas já morreram por causa de doenças associadas à AIDS desde o início da epidemia (UNAIDS, 2016). No Brasil, estima-se que em 2015 cerca de 830 mil pessoas (incidência de 0,04%) estavam infectados com o HIV (prevalência de 0,6%) e aproximadamente 15 mil mortes relacionadas com a infecção pelo HIV neste mesmo ano.

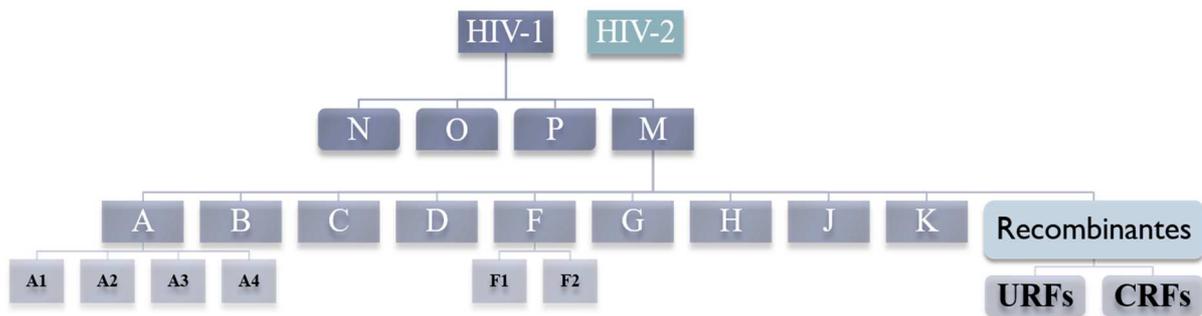
## 1.1 Classificação

O HIV é classificado em duas espécies diferentes: HIV-1 e HIV-2. Devido sua grande diversidade, o HIV-1 é classificado em quatro grupos filogeneticamente distintos: M, N, O e P. O grupo M é o mais disperso epidemiologicamente e possui grande diversidade genética sendo subclassificado em subtipos (A, B, C, D, F, G, H, J e K). Os subtipos A e F são ainda subclassificados em sub-subtipos A1, A2, A3, A4 e F1, F2, respectivamente (**Figura 2**) (VALLARI et al., 2011).

Além desta classificação deve-se levar em consideração a recombinação de diferentes subtipos entre si e entre os próprios grupos, formando vírus recombinantes. Por exemplo, vírus do subtipo B e do subtipo F podem recombinar seus genomas entre si gerando vírus recombinantes BF. Um vírus recombinante que é identificado em um único indivíduo ou em

um pequeno grupo de pessoas com relação epidemiológica é denominado Forma Recombinante Única (“*Unique Recombinant Form*” – URF). Entretanto, quando os vírus recombinantes são identificados com mesmo padrão de recombinação, através do sequenciamento do genoma completo, em pelos menos três indivíduos não relacionados epidemiologicamente são designados como Formas Recombinantes Circulantes (“*Circulating Recombinant Forms*” – CRFs).

**Figura 2** – Esquema da classificação do HIV-1 em grupos (N, O, P e M), subgrupos do grupo M (A-D, F-H, J-K), subsubgrupos do subgrupo A (A1- 4) e F (F1 - 2) e as URFs e CRFs.



Fonte: Próprio autor.

Para a classificação de uma CRF é necessário o sequenciamento de pelo menos três genomas completos ou dois genomas completos e um genoma parcial de vírus oriundos de indivíduos não relacionados epidemiologicamente e que estes genomas possuam um mesmo padrão de recombinação. A identificação de uma nova CRF é feita a partir de um número de identificação e das letras dos subtipos que lhe deram origem (por exemplo, CRF12\_BF), porém se três ou mais subtipos estão presentes, é utilizado *cpx* (*complex*) (ROBERTSON, 2000).

## 1.2 Epidemiologia molecular do HIV-1

As variantes genéticas do HIV estão dispersas heterogeneamente no mundo. No entanto, há variantes que predominam em certas regiões. Por exemplo, os subtipos A, C e D são os principais encontrados na África enquanto na Europa, Estados Unidos e Austrália predomina o subtipo B. Já o subtipo C é predominante na África do Sul, Etiópia e Índia e a CRF01\_AE no sudoeste da Ásia (PESSÔA et al., 2016).

A prevalência das formas recombinantes tem aumentado e estima-se que as URFs e CRFs são responsáveis por 20% das infecções pelo HIV no mundo (SIMONETTI et al., 2014). A CRF\_BF é a primeira CRF originada nas Américas, provavelmente a mais antiga

originada fora da África (CARMONA et al., 2002) e há fortes evidências que estes recombinantes tenham como ancestrais vírus F brasileiros (THOMSON et al., 2004). Estes recombinantes foram inicialmente isolados e caracterizados na Argentina (circulam neste país desde meados dos anos 80) (GOMEZ CARRILLO et al., 2002) e, mais tardiamente, foram descritos no Brasil, em 1994 (SABINO et al., 1994). Estima-se que a origem dos recombinantes BF no país ocorreu entre 1984 e 1991 (LEAL et al., 2008a).

Segundo dados de sequências de HIV depositadas no banco de dados do Laboratório Nacional de Los Alamos (LA), na América do Sul o subtipo B é predominante (66,6%), seguido do subtipo C (9,3%) e do subtipo F (7,9%). Já os recombinantes BF representam 12,4% das sequências de HIV recombinantes da América do Sul depositadas neste banco de dados, superando em número seu parental F. Considerando apenas as sequências de genomas recombinantes depositadas, os genomas de recombinantes BF chegam a 75,8% das sequências da América do Sul, região onde as primeiras CRFs\_BF foram descritas.

No Brasil, os subtipos B, C e F são predominantes porém, as diferentes regiões geográficas apresentam padrões distintos de distribuição destes subgrupos. Um aumento de subtipos não B, principalmente URFs\_BF também tem sido observado no Brasil (SANABANI et al., 2006; SANABANI et al., 2011; SANABANI et al., 2013). Um recente estudo de Pessôa e colaboradores mostrou que, em doadores de sangue de São Paulo, Rio de Janeiro, Minas Gerais e Recife, infectados com HIV-1, a prevalência dos subtipos é de 65% para o subtipo B, 28% para formas recombinantes (sendo 23,6% recombinantes BF), 3,9% para o C, 2% para o D e 1,5% para o F1 (PESSÔA et al., 2016). Um outro estudo de sequenciamento de nova geração (*Next Generation Sequencing* – NGS) em 2014 mostrou uma grande diversidade nos mosaicos estruturais de recombinantes BF1 e uma alta porcentagem de novos recombinantes CRF71\_BF1 e outros recombinantes BF1 entre doadores de sangue em Recife com altas taxas de mutações transmitidas de resistências à drogas (BRUSELLES et al., 2009). Outro estudo, conduzido por De Sá Filho e colaboradores, detectou uma taxa de aproximadamente 24% de CRF28 e CRF29 entre indivíduos infectados na cidade de Santos, estado de São Paulo, evidenciando o crescimento expressivo destes recombinantes no país (DE SA-FILHO et al., 2008).

Atualmente a prevalência de recombinantes BF tem aumentado nos países da América Latina e tem se tornado a forma recombinante mais comum, levando ao desaparecimento do subtipo parental F em algumas áreas (GOMEZ CARRILLO et al., 2002). A alta propensão dos subtipos B e F recombinarem pode ser explicada pelo um maior *fitness* viral comparado com os parentais, uma vez que é transmitido com uma maior frequência (AULICINO et al.,

2007). No entanto, tem se discutido a possibilidade de que a principal força evolutiva que tem levado ao declínio de parentais F e sua possível substituição pelo recombinantes BF pode ser atribuída a eventos estocásticos. Estes eventos seriam facilitados pelo baixo número da população F do que um aumento no *fitness* viral ou pressão seletiva em si (PESSÔA et al., 2015).

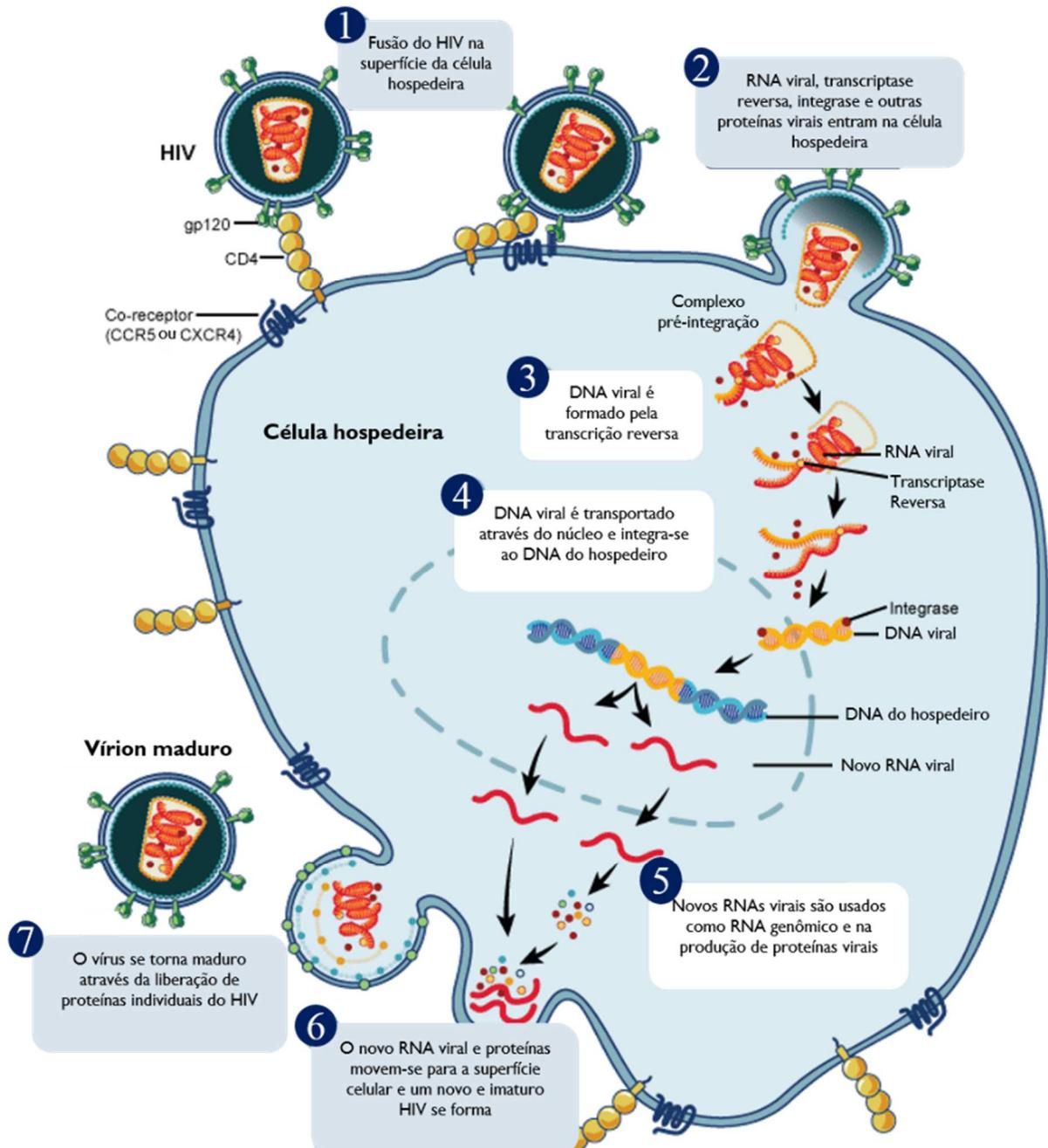
### 1.3 Ciclo de replicação

O ciclo de replicação do HIV, representado na **Figura 3**, pode ser dividido nas etapas de ligação, transcrição reversa, integração, transcrição, tradução, montagem e brotamento.

Inicialmente, a partícula viral interage com a célula hospedeira através da interação da proteína viral gp120 com receptor CD4 da célula hospedeira. Esta interação também acontece com um coreceptor de quimiocina específico, geralmente CCR5 ou CXCR4. Estas interações iniciam uma série de mudanças conformacionais que resulta na fusão do vírus com a membrana celular. Esta fusão leva a liberação do genoma viral, transcriptase reversa, integrase e outras proteínas virais para o interior da célula hospedeira. A esta fase inicial, dá-se o nome de fase de ligação e a mesma pode ser inibida pelos inibidores de fusão ou pelos inibidores de entrada. Na segunda etapa ocorre a transcrição reversa do genoma viral pela enzima transcriptase reversa, gerando uma DNA viral dupla fita (HU; TEMIN, 1990). Nesta etapa, podem ocorrer eventos de recombinação gerando vírus híbridos: com fragmento de genomas de vírus diferentes. Esta etapa pode ser inibida pelos inibidores da transcriptase reversa. Este DNA dupla fita é então direcionado para o núcleo e integrado ao genoma da célula hospedeira na etapa chamada integração. Esta etapa é realizada pela enzima viral integrase, que pode ser bloqueada pelos inibidores da integrase. Uma vez integrado ao genoma do hospedeiro, este DNA viral passa a ser chamado de DNA proviral.

A etapa da transcrição é mediada pela maquinaria da própria célula hospedeira. A enzima RNA polimerase II utiliza o DNA proviral como molde para a transcrição. Na fase da tradução os RNAs mensageiros são levados ao citoplasma e são traduzidos pela maquinaria transcricional do hospedeiro produzindo então proteínas e enzimas virais. Estas serão montadas juntamente com o genoma viral e direcionadas para uma região próxima da membrana plasmática. A protease viral cliva os polipeptídios em proteínas virais antes da maturação viral. Os inibidores da protease inibem esta fase do ciclo viral. Durante a fase de brotamento, os vírus são liberados da célula e podem infectar novas células, continuando o ciclo de infecção (ZHANG, 2007).

**Figura 3** – Ciclo de replicação do HIV.



Fonte: Disponível em [http://www.alyvea.com/micro/OSC\\_Microbio\\_06\\_02\\_hiv.jpg](http://www.alyvea.com/micro/OSC_Microbio_06_02_hiv.jpg)

#### 1.4 Recombinação e diversidade genética no HIV-1

Alguns fatores estão relacionados com a alta diversidade genética do HIV: a ausência da atividade corretora da transcriptase reversa viral ( $3 \times 10^{-5}$  mutações/nucleotídeo/ciclo de replicação), a alta taxa de replicação viral ( $10.3 \times 10^9$  partículas/dia) e o processo de recombinação durante a transcrição reversa (MELO; JAMAL; ZANOTTO, 2012). No

entanto, a taxa de mutação do HIV é similar a outros vírus de RNA sendo a diversidade genética encontrada nestes vírus atribuída em sua maior parte a sua elevada capacidade de recombinação em populações onde há mais de um subtipo de HIV circulante (FERNÁNDEZ-GARCÍA et al., 2010). A frequência de recombinação no HIV-1 é de  $2.4 \times 10^{-4}$ /nucleotídeo/ciclo de replicação, o equivalente a 2-3 eventos de recombinação por genoma e por ciclo de replicação, e ocorre ao longo de todo o genoma (JETZT et al., 2000).

O processo de recombinação contribui de maneira mais efetiva para a diversidade do HIV do que mutações pontuais. A recombinação pode ocorrer quando uma mesma célula é infectada por vírus diferentes. Na etapa da transcrição reversa, a transcriptase reversa pode, no momento da transcrição, mesclar entre fitas dos vírus diferentes em um processo chamado “jumping”. Neste caso há a formação de uma fita híbrida. Caso os vírus sejam pertencentes ao mesmo grupo o processo é chamado recombinação intrasubtipo. Porém, caso os vírus sejam pertencentes a grupos diferentes dá-se o nome de recombinação intersubtipo. Logo, a recombinação intersubtipo é um processo que ocorre principalmente em regiões nas quais há a circulação de subtipos virais diferentes uma vez o indivíduo necessariamente tem que estar coinfectado com diferentes subtipos virais.

Há um aumento considerável na descrição de URFs e CRFs nos últimos anos e este número tende a ser crescente uma vez que a distribuição geográfica dos subtipos é um processo dinâmico, tornando os eventos de recombinação entre as variantes de HIV-1 cada vez mais provável (GLAZIOU et al., 2015). Atualmente 88 CRFs e inúmeras URFs foram descritas (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). Desde a descrição da primeira CRF\_BF (CRF12) na Argentina em 2001 (CARR et al., 2001), diversas outras foram descritas posteriormente em diferentes partes do mundo. Das quatorze CRFs\_BF caracterizadas atualmente, duas foram identificadas na Argentina e em outros países da América do Sul (CRF12\_BF e CRF17\_BF), a CRF44\_BF no Chile, a CRF38\_BF no Uruguai, oito no Brasil (CRFs 28, 29, 39, 40, 46, 70, 71 e 72) e duas na Europa (CRF42 em Luxemburgo e CRF47 na Espanha). Os padrões de recombinação destas CRFs encontram-se listados no **Apêndice B**.

### 1.5 A alça V3 da proteína gp120

A alça do domínio V3 da proteína gp120 tem um importante papel na interação com o receptor e o coreceptor, processo crucial na entrada e o tropismo viral, além de ser um importante alvo para a neutralização com anticorpos (CUEVAS et al., 2010; VILLANOVA, 2010; MELO; JAMAL; ZANOTTO, 2012). Os aminoácidos presentes na alça do domínio V3 influenciam na utilização de determinado coreceptor. Portanto, o HIV-1 pode ser classificado de acordo com o coreceptor utilizado para a entrada viral. Aqueles que usam o coreceptor CCR5 são chamados de vírus CCR5-trópicos ou simplesmente vírus R5 e os que utilizam CXCR4 são classificados como vírus CXCR4-trópicos ou vírus X4. Há ainda aqueles que podem utilizar ambos receptores, os vírus R5X4 ou duplo-trópicos. Tem sido descrito na literatura que os vírus R5 são mais transmissíveis que vírus X4 e estão associados à rápida progressão da doença (CUEVAS et al., 2010). Além da associação com a progressão e estágio da doença, a predição da utilização do coreceptor é importante para o manejo clínico de fármacos antagonistas do coreceptor CCR5 pois pacientes infectados com vírus que podem utilizar o coreceptor CXCR4 não responderão a estes fármacos. Nesta região da alça V3 é encontrado um motivo estrutural tetrapeptídico altamente conservado nas posições 312-315 (em relação ao HXB2). Os tetrapeptídeos mais comuns são GPGR e GPGQ. As cepas brasileiras do subtipos B apresentam geralmente o motivo GWGR, que tornou-se uma assinatura característica destas cepas.

### 1.6 Resistência aos antirretrovirais

Desde a descoberta do HIV na década de 80, diversos esforços têm sido feitos em busca de uma terapia eficaz no combate ao vírus. Em 1986 a Zidovudina (AZT) foi aprovada para o tratamento contra o HIV e desde então outros fármacos foram introduzidos na terapia antirretroviral (ARV) para o HIV possibilitando o controle da doença e uma melhor qualidade de vida aos pacientes infectados. No entanto, apesar de inúmeros esforços científicos, ainda não foi descoberto um tratamento capaz de curar a infecção erradicando o vírus do paciente infectado que tenha avançado em ensaios clínicos.

Os ARVs são classificados de acordo com o mecanismo de ação. As principais classes de ARVs são: os Inibidores nucleosídicos da transcriptase reversa (*Nucleoside Reverse-Transcriptase Inhibitors* – NRTIs); Inibidores não nucleosídicos da transcriptase reversa (*Non-Nucleoside Reverse Transcriptase Inhibitors* – NNRTIs); Inibidores da protease

(*Protease Inhibitors* – PIs); Inibidores da Integrase (*Integrase Inhibitors* – INIs) e os Inibidores de fusão.

A resistência aos ARVs é oriunda de mutações específicas em determinadas regiões do genoma que podem ser resultantes do processo de mutação pontual ou do processo de recombinação (ARTS; HAZUDA, 2012) e alguns estudos mostram a relação entre a diversidade genética e as implicações clínicas como revisado por Santoro em 2013 (SANTORO; PERNO, 2013).

As mutações encontradas na região da protease e da integrase são classificadas em mutações principais (*Major*) e mutações acessórias. As mutações principais são, majoritariamente, mutações não polimórficas de resistência às drogas que tem uma grande contribuição na redução da susceptibilidade a uma ou mais drogas. Já as mutações classificadas como acessórias são mutações não polimórficas ou minimamente polimórficas que contribuem para a diminuição da susceptibilidade a determinada droga desde que encontrada em combinação com uma mutação principal. Algumas mutações pouco usuais ou mutações que ainda não foram bem caracterizadas encontradas nas posições de mutações principais também podem ser incluídas nesta lista.

A organização mundial da saúde (OMS) estima que em 2015 cerca de 46% das pessoas infectadas com HIV estavam em tratamento com ARVs no mundo, o dobro quando comparado a 2010 ([www.who.org](http://www.who.org)). O aumento das taxas de tratamento é positivo pois mostra que pacientes infectados com o HIV tem cada vez mais acesso ao tratamento. No entanto, considerando que os ARVs são um potente meio de pressão seletiva no genoma do HIV, deve-se levar em consideração um provável aumento da prevalência de variantes resistentes em função do aumento de pessoas sendo tratadas. Considerando estes aspectos, as sequências coletadas neste trabalho serão submetidas à análise de susceptibilidade viral aos ARVs.

## 2 JUSTIFICATIVA

O HIV-1 é caracterizado pela sua grande diversidade genética e esta é uma das características mais importante deste vírus e da pandemia da AIDS como um todo. Sendo atribuído a esta grande variedade a ampla disseminação e adaptação destes vírus, bem como a capacidade de escape da resposta imune do hospedeiro. Todas estas características resultam em dificuldades no desenvolvimento de vacinas e drogas capazes de erradicar o vírus.

Esta diversidade pode ser observada pela ampla classificação e subclassificação do HIV. Por exemplo, atualmente 88 CRFs e inúmeras URFs foram descritas (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>). Destas CRFs, 14 são recombinantes BF (15.9%) representando os duplo-recombinantes com o maior número de CRFs descritas. A primeira CRF\_BF foi descrita na Argentina em 2001 (CARR et al., 2001) e desde então estes recombinantes BF têm sido descritos em diversos países, principalmente no Brasil onde já foram identificadas oito CRFs\_BF e várias URFs\_BF.

Esta diversidade de recombinantes BF pode implicar em uma vantagem adaptativa uma vez que estes, ao surgirem, devem superar as cepas parentais dentro do indivíduo coinfetado para que sejam transmitidos eficientemente (MELO; JAMAL; ZANOTTO, 2012). A importância dos recombinantes BF na epidemia do HIV-1 tem sido demonstrada por diversos autores e reforçada pelo aumento do número destas variantes na população infectada. Há algumas linhas de evidência que demonstram que os recombinantes BF podem exibir diferentes características biológicas. Primeiro, foi observado o aumento da taxa de crescimento dos recombinantes BF para a CRF12\_BF e para a CRF38\_BF na Argentina e Uruguai (BELLO et al., 2010; AULICINO et al., 2007), que é corroborado por este trabalho principalmente em virtude da expansão de URFs BF no Brasil. A outra é evidência foi mostrada por um estudo publicado em 2012. Neste estudo os autores mostraram que o *Vpu* recombinante BF estava associado ao aumento na produção de partículas virais em ensaios celulares quando comparados com o subtipo B (DE CANDIA et al., 2010).

Portanto, estas características podem estar associadas ao aumento da dispersão destes recombinantes e a uma possível vantagem adaptativa em comparação com as cepas parentais e com outros recombinantes. Além disso, as consequências destas variações genótípicas refletem em diversos aspectos como, eficácia de testes de diagnóstico, monitoramento da doença, resposta imune, desfecho do tratamento e desenho de vacinas (HAMEELAR, 2013). Portanto, conhecer as variações genéticas do HIV se torna um ponto crítico para o desenvolvimento de drogas e para novas vacinas efetivas.

O número de genomas de HIV-1 depositados em banco de dados é crescente, inclusive para recombinantes BF. Porém, até o momento, não há trabalhos publicados que analisem estas sequências em conjunto com intuito de caracterizar os recombinantes BF circulantes. O objetivo deste trabalho foi delinear um panorama geral da diversidade dos recombinantes BF em todo o mundo. Com a expansão de genomas recombinantes HIV-1 BF e indícios que estes recombinantes tenham adquirido características que propiciem vantagem adaptativa em relação aos demais torna-se evidente a necessidade de uma caracterização dos mesmos com o objetivo de entender melhor estes recombinantes no contexto da epidemiologia do HIV-1. Esta caracterização poderá contribuir para uma melhor compreensão da epidemiologia do HIV e as possíveis interferências da expansão destes recombinantes no desenvolvimento de vacinas, seu impacto no desenvolvimento de novas drogas, fenótipo viral, patogenicidade, resposta a drogas já utilizadas e na epidemiologia da infecção do HIV-1 como um todo.

### **3 OBJETIVOS**

#### **Geral**

Caracterizar os isolados virais recombinantes entre os subtipos B e F circulantes em diversas regiões geográficas.

#### **Específicos**

1. Catalogar sequências e analisar informações associadas dos genomas completos do HIV-1 recombinantes BF publicados na literatura;
2. Investigar as relações filogenéticas dos recombinantes entre si e entre as cepas parentais;
3. Identificar e comparar os padrões de recombinação entre os genomas completos de recombinantes BF do HIV-1;
4. Investigar a existência de pontos preferenciais para ocorrência de recombinação genética (“hotspots”) nos genomas recombinantes BF do HIV-1;
5. Identificar assinaturas nucleotídicas conservadas nestes recombinantes;
6. Avaliar a presença de mutações associadas à resistência aos antirretrovirais nestes recombinantes;
7. Predizer o comportamento fenotípico dos recombinantes BF quanto a utilização preferencial do receptor de quimiocina na entrada viral.

## 4 MATERIAIS E MÉTODOS

### 4.1 Coleta de dados

As sequências de genomas completos de recombinantes BF do HIV-1 de todo o mundo foram coletadas do banco de dados de sequências de HIV do Laboratório Nacional de Los Alamos (LA) e do banco de dados *Nucleotide* do NCBI. Apenas genomas recombinantes BF com tamanho superior a 7000 pb foram considerados genomas completos. Este parâmetro é utilizado pelo banco LA para definir genoma completo (<http://www.hiv.lanl.gov/content/sequence/HIV/FAQ.html>). Estas sequências foram coletadas em arquivos no formato FASTA e as informações destes dois bancos de dados foram cruzadas juntamente com dados de artigos que descrevem a sequência para compor o banco de dados deste estudo.

Dentre as informações coletadas, quando disponíveis para dada sequência, estão o número de acesso da sequência no GenBank; identificação do paciente no LA; tamanho da sequência; artigo no qual a sequência foi descrita; subtipo informado no NCBI, em LA e nos artigos; país e cidade de coleta da amostra; gênero e idade do paciente; contagem de TCD4, TCD8 e carga viral; modo de transmissão; grupo de risco; ano de infecção, de diagnóstico e de coleta de amostra; estado clínico do paciente; utilização de coreceptor e fenótipo viral; e utilização de tratamento pelo paciente.

Estas informações coletadas foram utilizadas para confeccionar um banco de dados no Excel, programa implementado no pacote Office 2013. As informações inseridas no mesmo foram checadas em busca de erros ou divergência de dados.

### 4.2 Alinhamento e edição

Sequências de todos os subgrupos M do HIV-1 foram coletadas no LA (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) utilizando os filtros “subtype reference” para a opção tipo de alinhamento, “HIV-1/SIVcpz” para organismo, “genome” para a região, “M group without recombinants (A-K)” para subtipo, “DNA” para DNA/Proteína, “2010” para ano e “FASTA” para formato. Utilizando estes filtros, foram obtidas 39 sequências dos subtipos: A1 (DQ676872; AB253421; AB253429), A2 (AF286238; GU201516; AF286237), B (K03455; AY423387; AY173951; AY331295), C (U52953; U46016; AF067155; AY772699), D (K03454; AY371157; AY253311; U88824), F1

(AF077336; AF005494; AF075703; AJ249238), F2 (AY371158; AJ249236; AJ249237; AF377956), G (AF084936; AF061641; U88826; AY612637), H (AF190127; AF190128; AF005496; FJ711703), J (EF614151; GU237072; AF082394) e K (AJ249235, AJ249239), totalizando 39 sequências. O genoma do isolado ANT70 (número de acesso no GeneBank: L20587.1), pertencente ao grupo O, foi utilizado como grupo externo.

As sequências GU595155, GU595159, GU595161, GU595162, KJ849757 e KJ849772 foram excluídas especificamente desta análise por apresentarem diversas bases degeneradas. No entanto, estas sequências foram mantidas nas demais análises em virtude da boa qualidade do fragmento utilizado como, por exemplo, a região *pol* para a análise de resistência a drogas e a região da GP120 para a predição da utilização de coreceptor. O alinhamento final é composto por 246 sequências recombinantes BF, mais 39 sequências referência e a sequência do grupo externo, totalizando 286 sequências. As sequências deste trabalho foram identificadas da seguinte maneira: número de acesso do GeneBank.código do país.ano de coleta.classificação. Já as sequências referência foram identificadas como subtipo.número de acesso do GeneBank.

O alinhamento foi realizado utilizando o algoritmo *muscle* (v3.8.31) disponível no programa SeaView (v. 4.5.4 para Windows) (GOUY; GUINDON; GASCUEL, 2010). A edição manual do alinhamento foi feita utilizando o programa BioEdit (HALL, 1999). Após a edição uma cópia do alinhamento foi utilizada para obter um alinhamento com o mínimo fragmento compartilhado entre todas sequências (5352 pb) que foi utilizado nas análises de reconstrução filogenética.

### 4.3 Reconstrução filogenética

O modelo de substituição de nucleotídeos mais apropriado para cada análise foi determinado pelo programa ModelTest 3.7 (POSADA; CRANDALL, 1998). O modelo escolhido foi GTR+G+I (General Time-Reversible + gamma-distribution + invariable sites), definido segundo os critérios de informação de Akaike (AIC). Os parâmetros obtidos para este modelo foram: frequências de base (freqA = 0.4497, freqC = 0.1654, freqG = 0.2134, freqT = 0.1715); taxas da matriz: R(a) [A-C] = 1.3365; R(b) [A-G] = 3.9979; R(c) [A-T] = 0.7110; R(d) [C-G] = 0.6810; R(e) [C-T] = 6.9523; R(f) [G-T] = 1.0000; proporção de sítios invariáveis (I) = 0.2483; e distribuição gamma = 0.6823.

O comportamento filogenético das sequências dos vírus recombinantes BF foram comparados entre si, entre os subtipos parentais e entre outros subtipos puros. A reconstrução

da história evolutiva destas sequências foi estimada através de análises filogenéticas utilizando os métodos “neighbor-joining” (NJ) e máxima verossimilhança (“Maximum Likelihood” - ML) utilizando o programa PAUP\* versão 4.0b (SWOFFORD, 2002). O critério utilizado para a busca heurística da árvore foi o algoritmo de troca entre vizinhos mais próximos NNI (*Nearest-Neighbor Interchange*) e a sustentação dos ramos da árvore foi calculada através do cálculo de “*bootstrap*” baseado em 1000 reamostragens. Os valores de *bootstrap* foram utilizados na árvore ML quando a topologia dos clados foram repetidas tanto na NJ quanto na ML. As árvores foram visualizadas e editadas utilizando o programa FigTree v 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

#### 4.4 Análises de recombinação

Os padrões de recombinação foram estimados utilizando jpHMM (*jumping profile Hidden Markov Model*), disponível em uma plataforma online ([http://jphmm.gobics.de/submission\\_hiv.html](http://jphmm.gobics.de/submission_hiv.html)). As sequências dos genomas completos foram submetidas a ferramenta online e os padrões de recombinação gerados a partir desta análise foram salvos. Esta ferramenta utiliza informações detalhadas dos polimorfismos da população parental, além de usar cepas parentais nas análises, e provê informações detalhadas da confiabilidade dos pontos de recombinações preditos.

O jpHMM utiliza algoritmos baseados no modelo oculto de Markov que comparam a sequência analisada com alinhamentos múltiplos de sequências de diferentes subtipos de HIV. A partir do alinhamento de cada subtipo é criado um modelo probabilístico com o qual a sequência será comparada. No entanto, diferentes fragmentos da sequência analisada são alinhados localmente aos diferentes subtipos, permitindo assim a identificação de recombinantes com maior facilidade e rapidez (MELO, 2011).

#### 4.5 Análise das características moleculares

As sequências consenso B, F1 e F2 utilizadas para comparação destas análises foram obtidas a partir do LA utilizando as opções “Web” para o filtro tipo de alinhamento, “HIV-1/SIVcpz” para o filtro organismo, “Genome” para o filtrar a região, “M group without recombinants” para o filtro subtipo; “DNA” para o filtro DNA/proteína; “2015” para o filtro ano e “FASTA” para o formato (<https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>). Esta busca retorna um

alinhamento de 1666 sequências de todos subtipos do grupo M. As sequências pertencentes aos subtipos B (n = 807), F1 (n = 27) e F2 (n = 6) foram separadas em diferentes arquivos e a sequência consenso de cada um destes grupos foi gerada no Geneious.

Para a análise de assinaturas nucleotídicas foram analisadas as posições 790 a 9417 (posição em relação a HXB2) de todas as sequências recombinantes BF do alinhamento juntamente com as sequências referência B e F já citadas anteriormente, excluindo as sequências referência dos demais subtipos. Esta análise foi realizada no programa Geneious, utilizando a opção “Annotate & Predict” e posteriormente a opção “Find variations/SNPs”.

#### **4.6 Análise de resistência aos antirretrovirais**

As análises de resistência à drogas foram realizadas utilizando o algoritmo de interpretação de resistência genotípica do Banco de Dados de Resistência a Drogas do HIV da Universidade de Stanford (HIVdb) versão 8.1.1 (atualizada em 23/09/2016) (<https://hivdb.stanford.edu/hivdb/by-sequences/>). Este algoritmo, implementado na plataforma online do banco de dados, permite a identificação de mutações que podem conferir resistência, em diferentes níveis de classificação, aos inibidores da protease, da transcriptase reversa e da integrase que são comumente utilizados. As análises são subdivididas em quatro classes de acordo com o alvo de atuação dos fármacos: transcriptase reversa (NRTIs e NNRTIs), protease (PIs) e integrase (INIs). Os fármacos avaliados para a classe dos NRTIs são Abacavir (ABC), Zidovudina (AZT), Estavudina (D4T), Didanosina (DDI), Emtricitabina (FTC), Lamivudina (3TC) e Tenofovir (TDF). Para a classe dos NNRTIs foram analisados os fármacos Efavirenz (EFV), Etravirina (ETV), Nevirapina (NVP) e Rilpivirina (RPV). O Atazanavir (ATV), Darunavir (DRV), Fosamprenavir (FPV), Indinavir (IDV), Lopinavir (LPV), Saquinavir (SQV) e o Tipranavir (TPV) foram os fármacos analisados para classe dos PIs enquanto o Dolutegravir (DTG), Elvitegravir (EVG) e o Raltegravir (RAL) foram os fármacos da classe dos INIs avaliados.

Nas análises do Stanford Database, o nível de resistência estimado para a droga é determinado pelo algoritmo através de um mecanismo que adiciona escores associados a cada mutação encontrada na sequência e em alguns casos a combinações de mutação. A presença de determinada mutação tem um valor fixo de escore para cada droga. Por exemplo, a presença da mutação L10F na protease soma 0 pontos de escore para o fármaco ATV enquanto para os fármacos FPV e NFV somam 15 pontos para cada. A partir dos escores individuais de cada mutação para determinado fármaco é calculado um escore total. Este valor

determina a qual dos cinco níveis de resistência a mutação está associada para cada droga analisada. Os níveis de resistência são classificados em susceptível (escore de 0-9), potencial (escore de 10-14), baixo (escore de 15-29), intermediário (escore de 30-59) e alto (escore maior ou igual a 60). Um nível de resistência classificado como susceptível indica que não há evidências de que a mutação (ou as mutações) presente na sequência leve a uma redução da susceptibilidade ao ARV analisado quando comparado como o vírus tipo selvagem. O nível de resistência considerado potencial pode indicar que a sequência pode conter mutações que indicam tratamento prévio com ARV ou a sequência pode conter mutações que propicie uma susceptibilidade reduzida *in vitro* ao fármaco avaliado apenas quando esta mutação estiver combinada com outras. Já o nível de resistência baixo sinaliza que o vírus do qual a sequência foi oriunda pode ter susceptibilidade reduzida ou ainda que os pacientes infectados com tais vírus podem apresentar uma resposta virológica subótima ao tratamento com ARV. O nível de resistência intermediário representa uma alta probabilidade de redução da atividade da droga e um alto nível de resistência indica que pacientes infectados com estes vírus tendem a apresentar uma resposta virológica muito baixa ou não apresentam resposta virológica àquela droga.

#### **4.7 Predição de uso do coreceptor**

Para a predição genotípica do coreceptor utilizado na entrada viral foi utilizada a ferramenta de bioinformática geno2pheno do Instituto de Informática Max Plank (<http://coreceptor.geno2pheno.org/index.php>). A sequência completa foi submetida na plataforma online para análise. Este algoritmo utiliza a região da alça V3 da proteína gp120 do envelope viral para predizer se o vírus correspondente é capaz de usar o coreceptor CXCR4 como coreceptor (vírus R5/X4 ou X4) ou não (vírus R5). Portanto, a classificação de vírus nesta plataforma como R5/X4 não distingue a capacidade do vírus utilizar apenas o coreceptor CXCR4 (X4) ou do vírus ser duplo-trópico (R5/X4), podendo utilizar ambos coreceptores no processo da entrada viral.

A plataforma recomenda que as sequências sejam analisadas em triplicatas e quando apenas uma sequência é analisada a taxa de falso positivo (FPR) deve ser de 20%. Este valor é utilizado no intuito de minimizar o número de falsos positivos (identificar vírus capazes de utilizar o coreceptor CXCR4 como R5). Esta taxa tende a propiciar uma maior confiabilidade no resultado para as sequências identificadas como provenientes de vírus capazes de utilizar

apenas o coreceptor CXCR5 e portanto, estes pacientes podem ser tratados com inibidores de CCR5 com uma margem de segurança maior.

#### **4.8 Análise estatística**

As análises estatísticas foram realizadas no GraphPad Prism.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Caracterização das sequências

As sequências foram coletadas no formato FASTA a partir do NCBI e do LA até junho de 2016 totalizando 252 sequências (**Apêndice A**) de genomas completos de isolados virais recombinantes BF do HIV-1.

É importante salientar a importância dos dados clinico-epidemiológicos para o estudo de sequências de HIV. Portanto, seria oportuno se os bancos de dados exigissem informações básicas dos pacientes dos quais as sequências foram geradas. Na maioria dos casos, o pouco de informação coletada sobre cada sequência só foi possível devido o cruzamento de dados encontrados em artigos, no NCBI e em LA. Ainda assim, informações importantes sobre os pacientes dos quais estas sequências foram isoladas nem sempre estavam disponíveis, como evidenciado na **Tabela 1**. Portanto, as discussões feitas a seguir serão embasadas nos dados disponíveis.

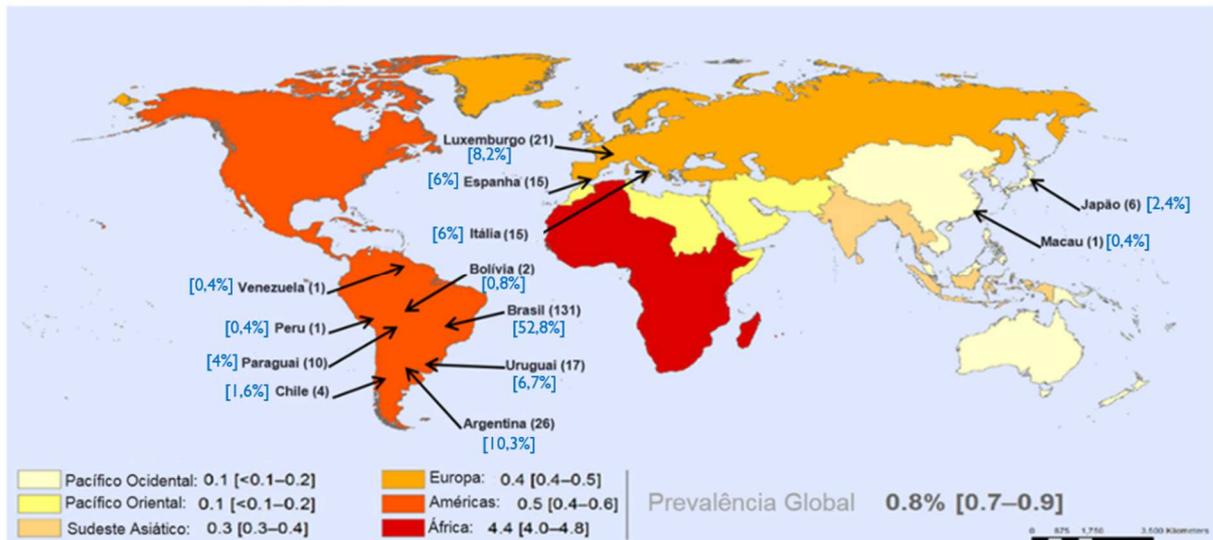
**Tabela 1**– Informações coletadas sobre as sequências.

DADOS	D (n)	%
País	252	100
Ano de coleta	250	99
Subtipo - LA	235	93
Subtipo - Artigo	210	83
Subtipo-NCBI	203	81
Cidade/Estado	141	56
Gênero	133	53
Grupo de Risco	114	45
Idade	93	37
Via de transmissão	74	29
Ano de diagnóstico	39	15
Estado clínico	35	14
CD4	29	12
Carga Viral	29	12
Coreceptor	22	9
Tratamento	21	8
Fenótipo	9	4
Ano de infecção	7	3
CD8	2	1

Fonte: Dados coletados pela autora. D: dado disponível.

As sequências coletadas são oriundas de treze países diferentes: 133 do Brasil (BR; 52,8%), 26 da Argentina (AR; 10,3%), 21 de Luxemburgo (LX; 8,2%), 17 do Uruguai (UR; 6,7%), 15 da Espanha (ES; 6%), 15 da Itália (IT; 6%), 10 do Paraguai (PR; 4%), 6 do Japão (JP; 2,4%), 4 da Chile (CH; 1,6%), 2 da Bolívia (BO; 0,8%) e 1 da Venezuela (VZ), 1 do Peru (PE) e 1 de Macau (MC) (0,4% cada) (**Figura 4**). Estes achados mostram a dispersão global destes recombinantes que inicialmente foram descritos na América do Sul.

**Figura 4** – Caracterização geográfica dos recombinantes BF do HIV-1 circulantes no mundo.



Fonte: Mapa adaptado da WHO, disponível em [http://www.who.int/gho/hiv/hiv\\_013.jpg?ua=1](http://www.who.int/gho/hiv/hiv_013.jpg?ua=1) Mapa da prevalência de infecções por HIV entre adultos de 15-49 anos (dados de 2015), com a indicação do número de sequências coletadas de acordo como país de coleta.

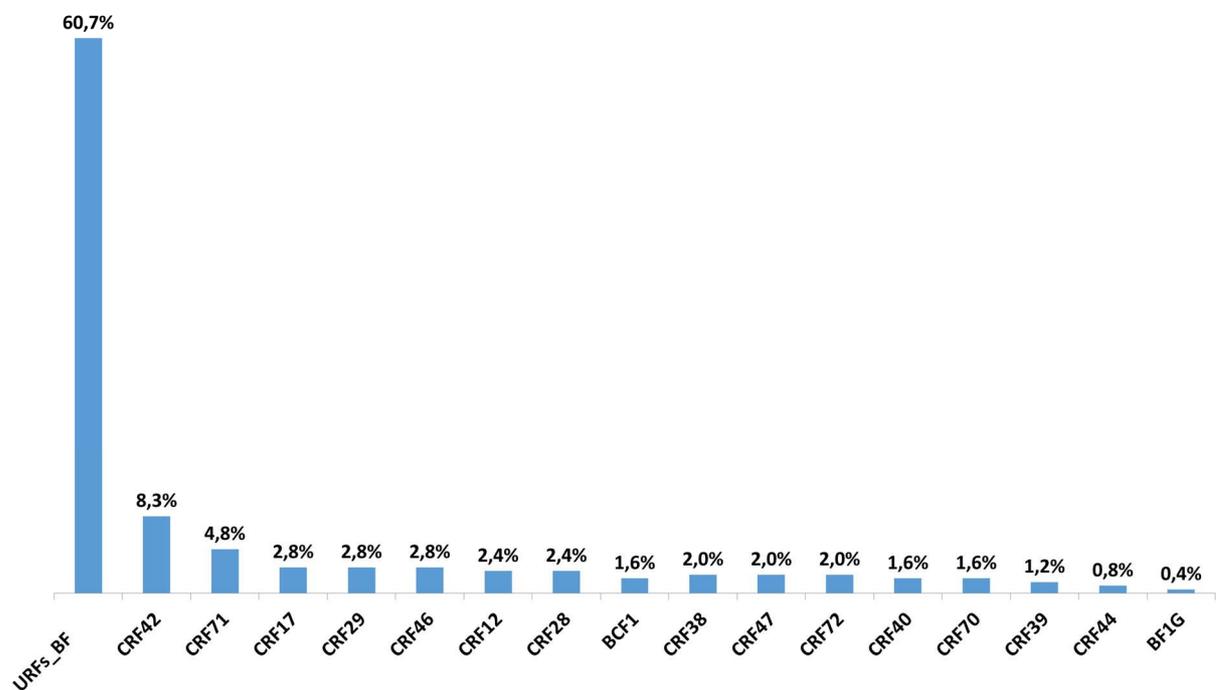
O tamanho médio das sequências foi de 8879 pb (variando de 7293 pb – 9770 pb). Considerando que 86% das sequências apresentaram tamanho entre 8500 e 9200 pb, pode-se perceber que a maioria das sequências possui a cobertura quase completa do genoma.

A mediana do ano de diagnóstico foi 1996, variando de 1985 a 2008. A sequência coletada do paciente diagnosticado em 1985 (ano de diagnóstico mais antigo), do sexo masculino e usuário de drogas injetáveis é a sequência JF804806.BR.2005.29. Esta sequência foi coletada em 2005 no Brasil e pertence a CRF29. Portanto, este recombinante pode ter circulado antes mesmo da descrição do primeiro recombinante neste país em 1992 (MELO; JAMAL; ZANOTTO, 2012).

A amostra mais antiga foi coletada em 1993 e a mais recente em 2014 (mediana: 2004). Já a mediana para a idade dos pacientes no momento da coleta foi de 31,5 anos e a maioria (67,7%; 63/93) dos pacientes pertencem a faixa etária entre 20 a 39 anos. Esta faixa coincide com a faixa maior incidência de infecção por HIV (WHO, 2014).

Quanto à classificação das sequências (**Gráfico 1**), a maioria é classificada como URFs\_BF (153/252; 60,7%). Outras 21 sequências estão classificadas como pertencentes à CRF42 (21/252; 8,3%), 12 sequências classificadas como CRF71 (12/252; 4,8%), 7 como CRF17 (7/252; 2,8%), 7 como CRF29, 7 como CRF46 (7/252; 2,8%), 3 como CRF39 (3/252; 1,2%) e 2 como CRF44 (2/252; 0,8%). Já as CRF12 e CRF28 tiveram 6 sequências cada (6/252; 2,4% cada), as CRF38, CRF47 e CRF72 com 5 sequências cada (5/252; 2%), enquanto a CRF40 e CRF70 tiveram 4 sequências (4/252; 1,6%). Além destas, outras quatro como URFs BCF<sub>1</sub> (4/252; 1,6%) e uma sequência foi classificada como uma URF BF<sub>1</sub>G (1/252; 0,4%). Portanto, considerando a grande quantidade de vírus recombinantes BF no Brasil e o fato que recombinantes BF com vírus parentais preditos como brasileiros foram identificados na Itália e Japão, pode-se observar que os recombinantes BF brasileiros tem um importante papel na pandemia global do HIV (BRUSELLES et al., 2009; SANABANI et al., 2010).

**Gráfico 1** – Classificação das sequências.



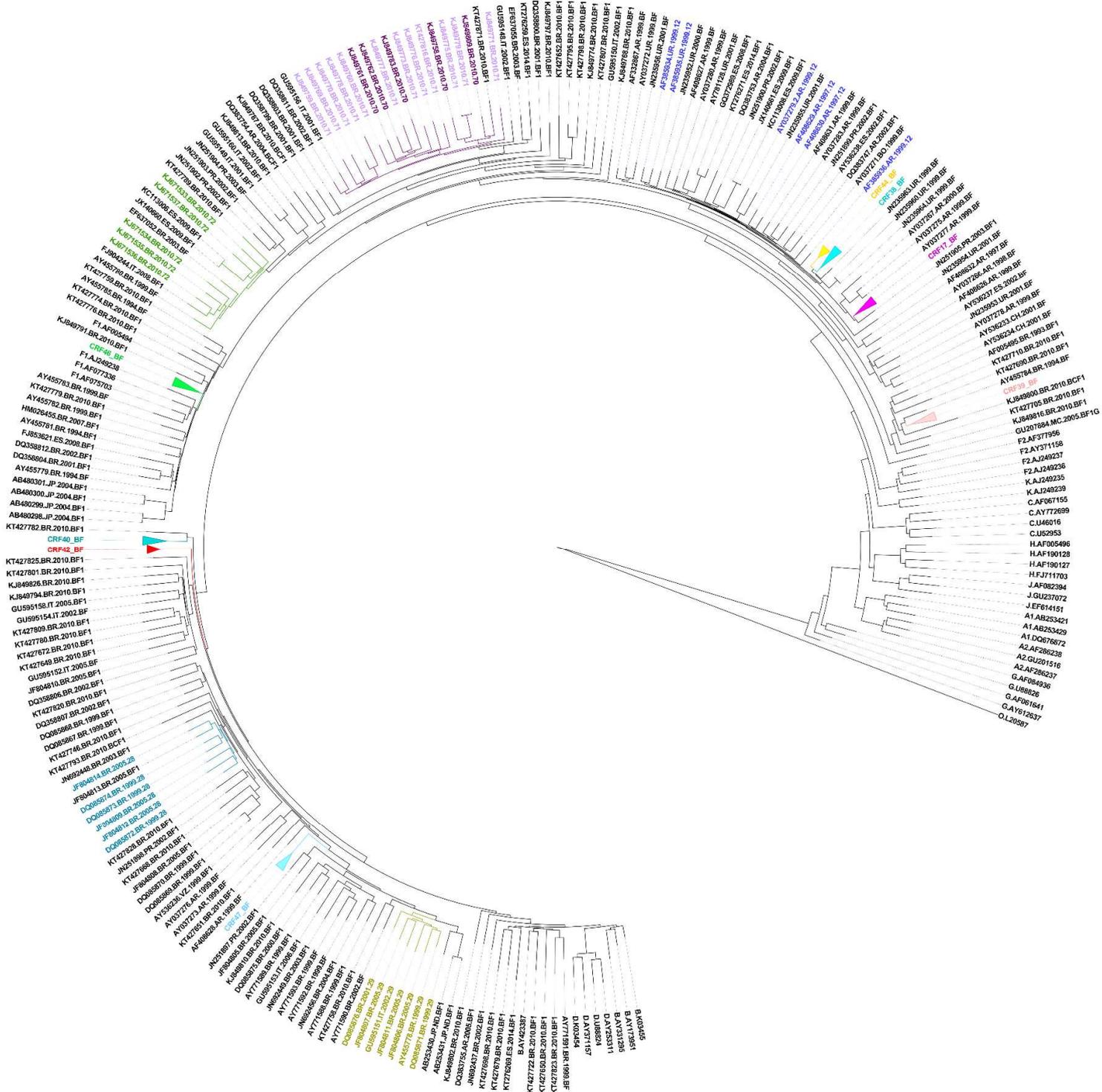
Fonte: Dados coletados pela autora.

## 5.2 Análise filogenética

A reconstrução da história evolutiva das sequências analisadas no presente estudo está representada na árvore ML (**Figura 5**) (A árvore com leiaute retangular está disponível no

material suplementar disponível no link [goo.gl/nyKuKZ](http://goo.gl/nyKuKZ)), enraizada pelo ponto médio e a escala representa a substituição de 0,2 nucleotídeos por sítio.

Figura 5 - Árvore filogenética ML com leiaute circular.



Fonte: Dados coletados pela autora. Cada cor identifica uma sequência CRF. A escala representa a substituição de 0,2 nucleotídeos por sítio.

A árvore filogenética construída utilizando o método NJ também está disponível no material suplementar ([goo.gl/nyKuKZ](http://goo.gl/nyKuKZ)). Esta árvore possui a topologia bastante similar com a árvore ML, o que fortalece a confiança nas relações topológicas observadas. Os valores de bootstrap acima de 70 de topologias que se repetiram tanto na árvore NJ quanto na árvore ML foram identificados nesta última árvore.

Analisando as sequências de CRFs\_BF observa-se que algumas delas formaram grupos únicos. A **CRF38** e a **CRF44** formaram, cada uma delas, um clado único com *bootstrap* de 100 (**Figura 6**).

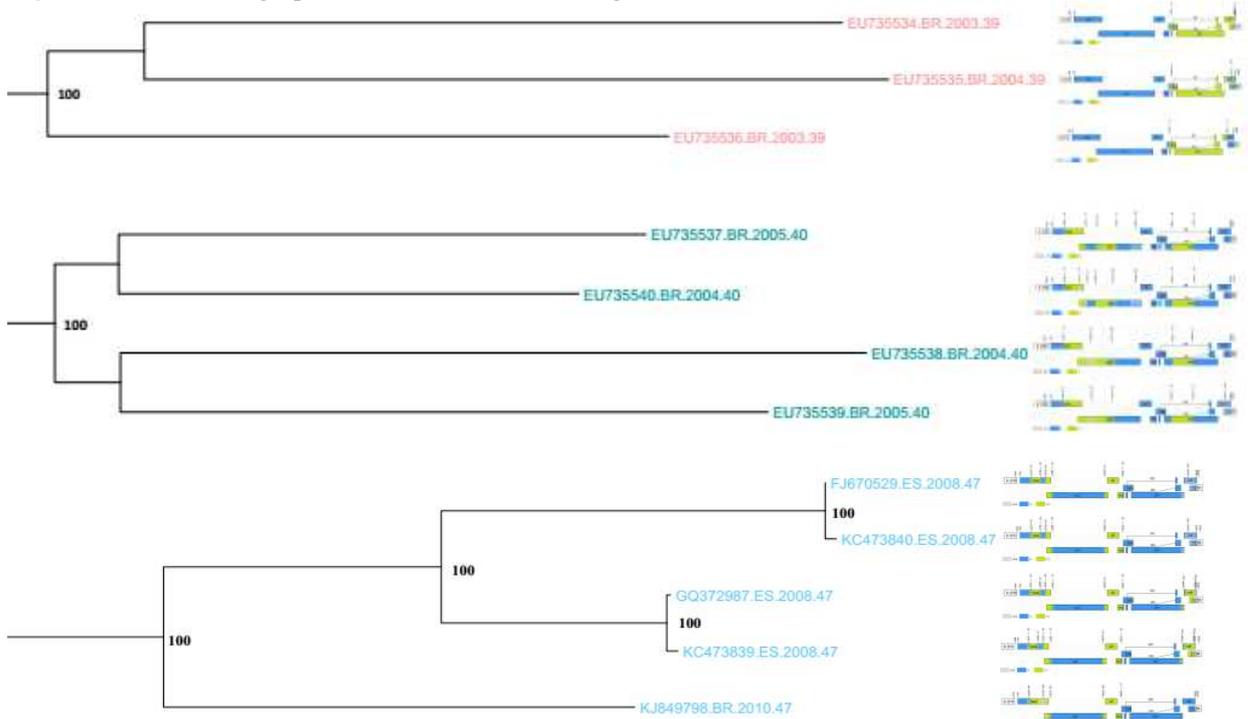
**Figura 6** - Cluster da árvore ML retangular contendo as CRFs 38 e 44.



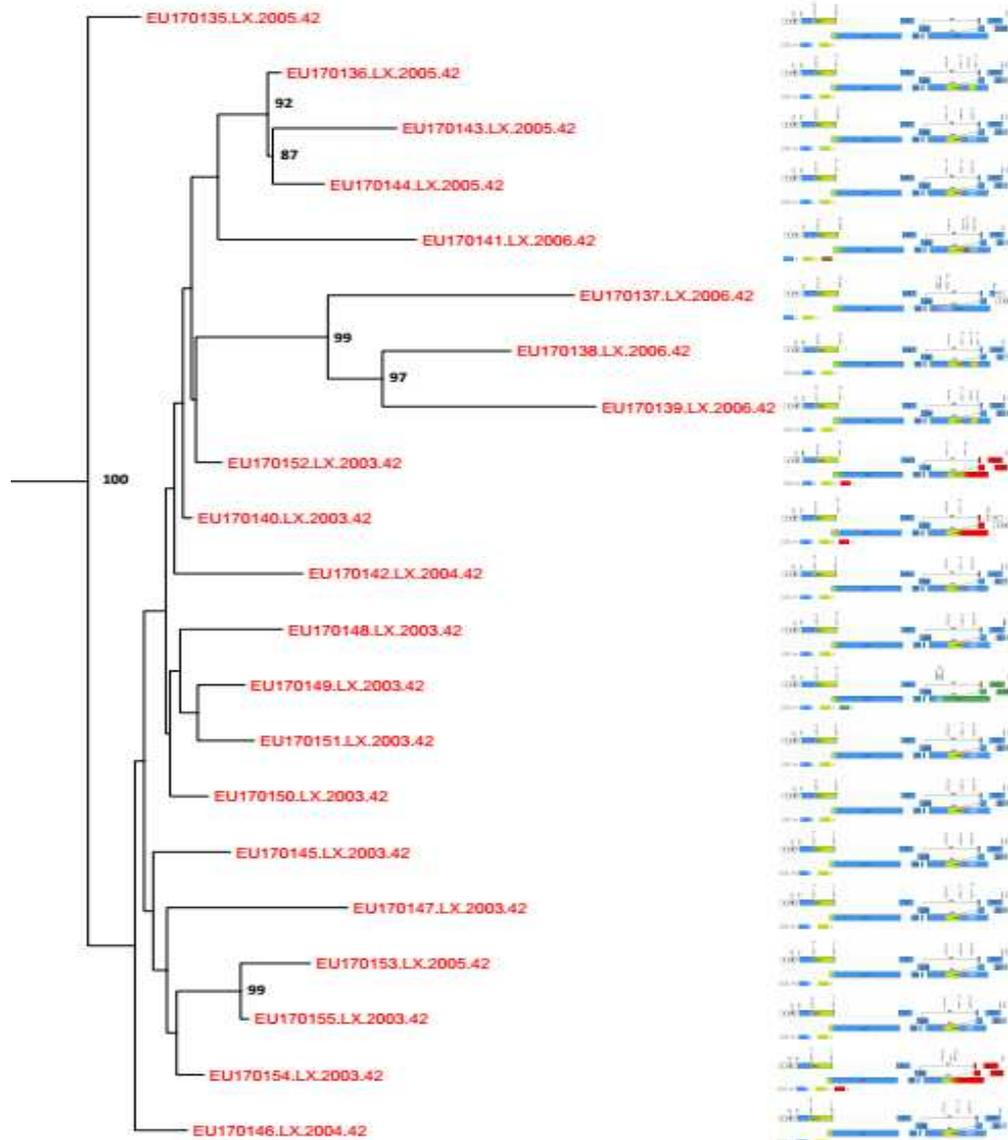
Fonte: Dados coletados pela autora.

As sequências pertencentes às **CRF39**, **CRF40**, **CRF42** e **CRF47** agruparam em clados únicos com valor de *bootstrap* 100 (**Figura 7**).

**Figura 7** - Diferentes agrupamentos da árvore ML retangular contendo as CRF 39, 40, 42 e 47.



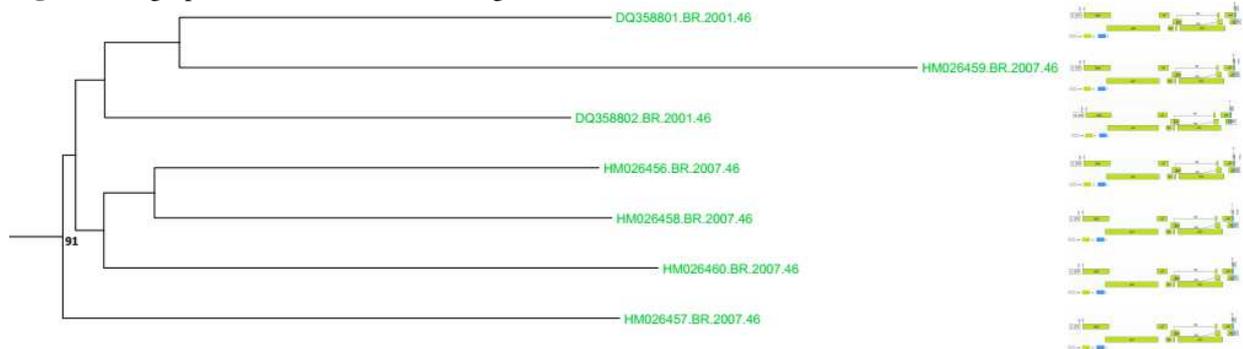
**Figura 7** – Continuação. Diferentes agrupamentos da árvore ML retangular contendo as CRF 39, 40, 42 e 47.



Fonte: Dados coletados pela autora.

Já as sequências da **CRF46** formaram um clado único com *bootstrap* 91 (**Figura 8**). No entanto, as sequências AY455781.BR.1994.BF1, HM026455.BR.2007.BF1, AY455782.BR.1999.BF1, KT427779.BR.2010.BF1, KT427774.BR.2010.BF1, AB480298.JP.2004.BF1, AB480299.JP.2004.BF1, AB480300.JP.2004.BF1 e AB480301.JP.2004.BF1 possuem padrão de recombinação similar ao padrão da CRF46 e estão difusas na árvore. Estas sequências relacionadas a CRF46 parecem ter sido originadas de eventos de recombinação diferentes (SANABANI et al., 2010). No entanto, como o critério para classificação de CRF é possuir o mesmo padrão de recombinação sugere-se a reclassificação destas sequências como CRF46.

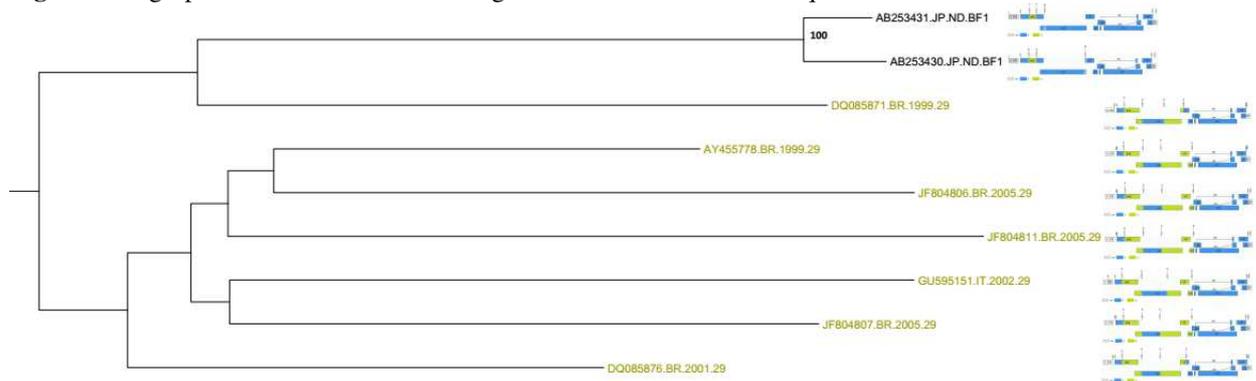
**Figura 8** - Agrupamento da árvore ML retangular contendo a CRF46.



Fonte: Dados coletados pela autora.

Considerando o agrupamento de todas as sequências da **CRF29** (**Figura 9**) pode-se observar que estas agruparam com outras duas sequências AB253431.JP.ND.BF1 e AB253430.JP.ND.BF1. No entanto, estas sequências possuem padrão de recombinação diferente da CRF29.

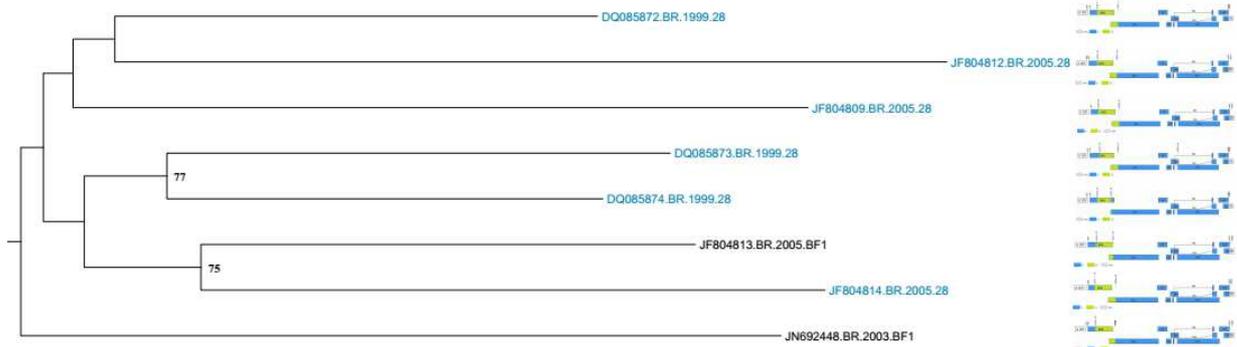
**Figura 9** - Agrupamento da árvore ML retangular contendo a CRF29 e sequências relacionadas.



Fonte: Dados coletados pela autora.

As sequências da **CRF28** (**Figura 10**) formaram um cluster com a sequência JF804813.BR.2005.BF1 que agrupou especificamente com a sequência JF804814.BR.2005.28 com valor de *bootstrap* de 75. Portanto sugere-se a reclassificação desta sequência relacionada para CRF28 uma vez que formaram um grupo monofilético com as sequências desta CRF.

**Figura 10** - Agrupamento da árvore ML retangular contendo a CRF28 e sequências relacionadas.



Fonte: Dados coletados pela autora.

O grupo que contém todas as sequências da **CRF72** (**Figura 11**) apresentou um valor de *bootstrap* de 95. As sequências destas CRF agruparam com outras sequências de padrão de recombinação similar: AY455780.BR.1999.BF, FJ904244.IT.2008.BF1, EF637052.BR.2003.BF, JX140660.ES.2009.BF1, KC113006.ES.2009.BF1, e KT427789.BR.2010.BF. Esta última sequência possui uma segunda inserção do subtipo B, quando comparada com o padrão da CRF72, na região 1760+-59/2014+-27. A sequência EF637052.BR.2003.BF já foi descrita anteriormente com o mesmo padrão de recombinação que a CRF72 (PESSOA et al., 2014) (THOMSON et al., 2004). Percebe-se portanto que o cluster que contém as sequências da CRF72 é heterogêneo e que, as demais sequências deste cluster devem ser reclassificadas como CRF72, excetuando a sequência KT427789.BR.2010.BF e FJ904244.IT.2008.BF. Esta última sequência tem um fragmento do subtipo B na região *env* maior que o encontrado na CRF72 que pode ter sido adquirido por um evento de recombinação adicional no ancestral desta sequência que possuía o mesmo padrão que a CRF72.

**Figura 11** - Agrupamento da árvore ML retangular contendo a CRF72 e sequências relacionadas.

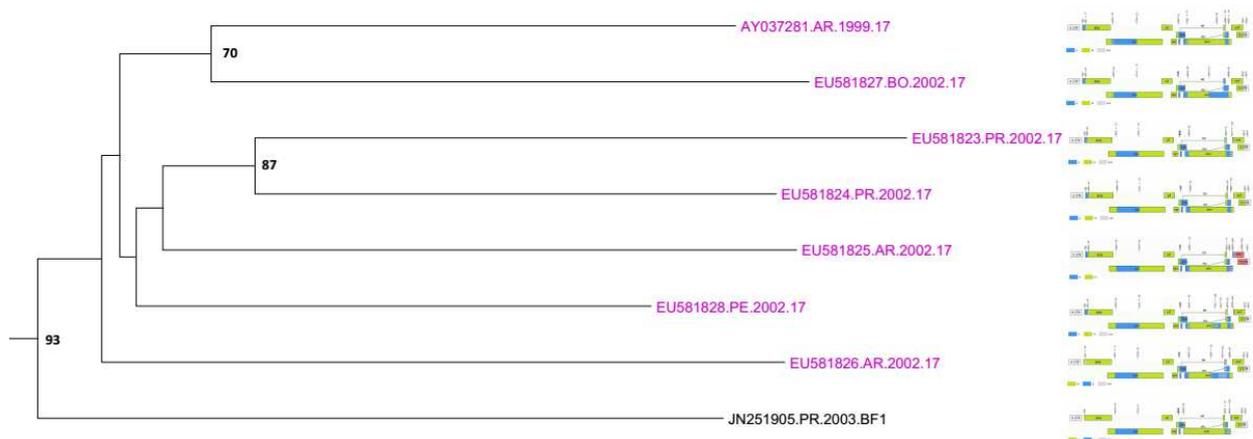


Fonte: Dados coletados pela autora.

Já as sequências classificadas como **CRF17** formaram um clado único (**Figura 12**). Sendo que a sequência JN251905.PR.2003.BF1 está relacionada com esta CRF (*bootstrap* de 93). Em virtude do alto valor de *bootstrap* com que estas sequências agruparam e do padrão de recombinação similar pode-se sugerir que a sequência JN251905.PR.2003.BF1 deve ser reclassificada como CRF17.

As sequências da **CRF70** e da **CRF71** formaram um grande grupo (*bootstrap* de 94) na qual estão misturadas em outros subgrupos juntamente com outras sequências que possuem padrão de recombinação similares a esta CRF (DQ358811.BR.2002.BF1, GU595156.IT.2001.BF1 e KT427871.BR.2010.BF1) (**Figura 13**). Porém uma análise das assinaturas nucleotídicas mostrou que, provavelmente, as URFs contidas neste grupo são mais próximas ou pertencem à CRF71. Como já discutido, a sequência DQ358811.BR.2002.BF1 tem o padrão de recombinação quase idêntico à CRF71\_BF (PESSOA et al., 2014).

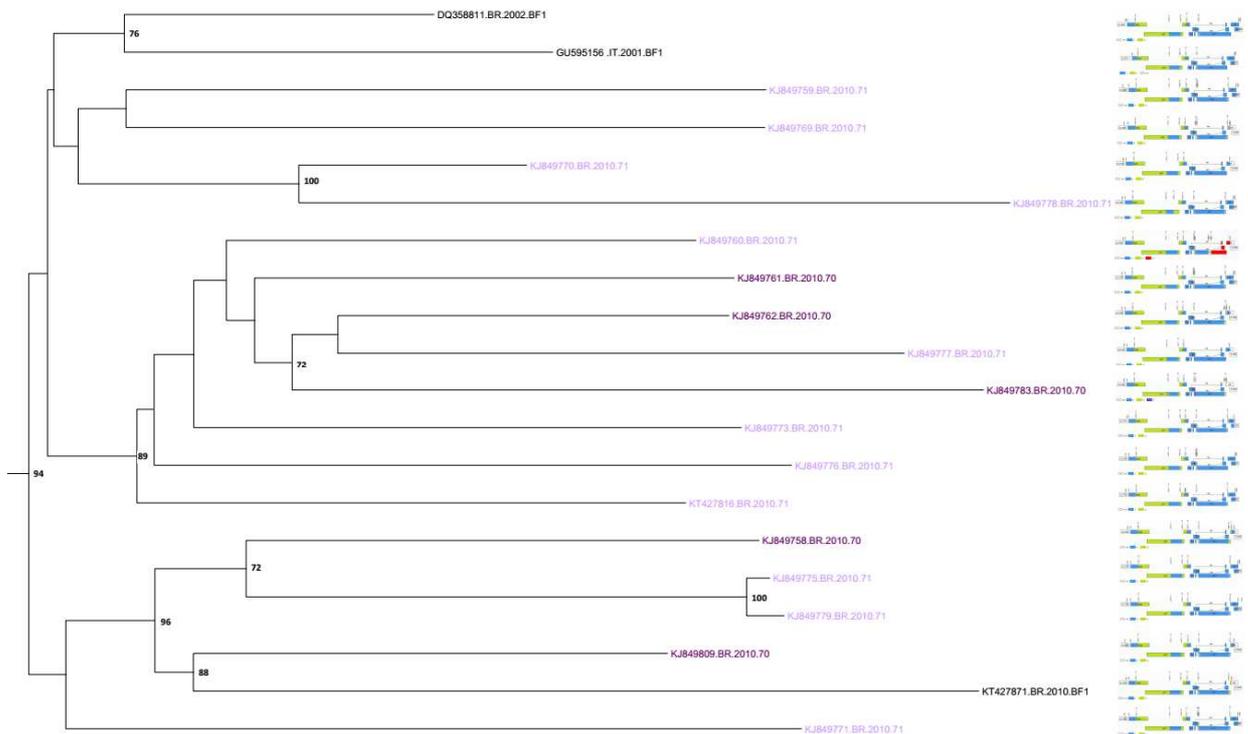
**Figura 12** - Agrupamento da árvore ML retangular contendo a CRF17 e sequências relacionadas.



Fonte: Dados coletados pela autora.

A CRFs\_BF 70 e 71 possuem padrão de recombinação bem similares. Os autores que descreveram esta CRF discutiram a possibilidade da CRF70 ter sido originada a partir de um evento de recombinação secundário da CRF71 e que elas tem evoluído independentemente desde então já que a distância entre elas é de aproximadamente 8% (PESSOA et al., 2014).

**Figura 13** - Agrupamento da árvore ML retangular contendo a CRF70, CRF71 e seqüências relacionadas.

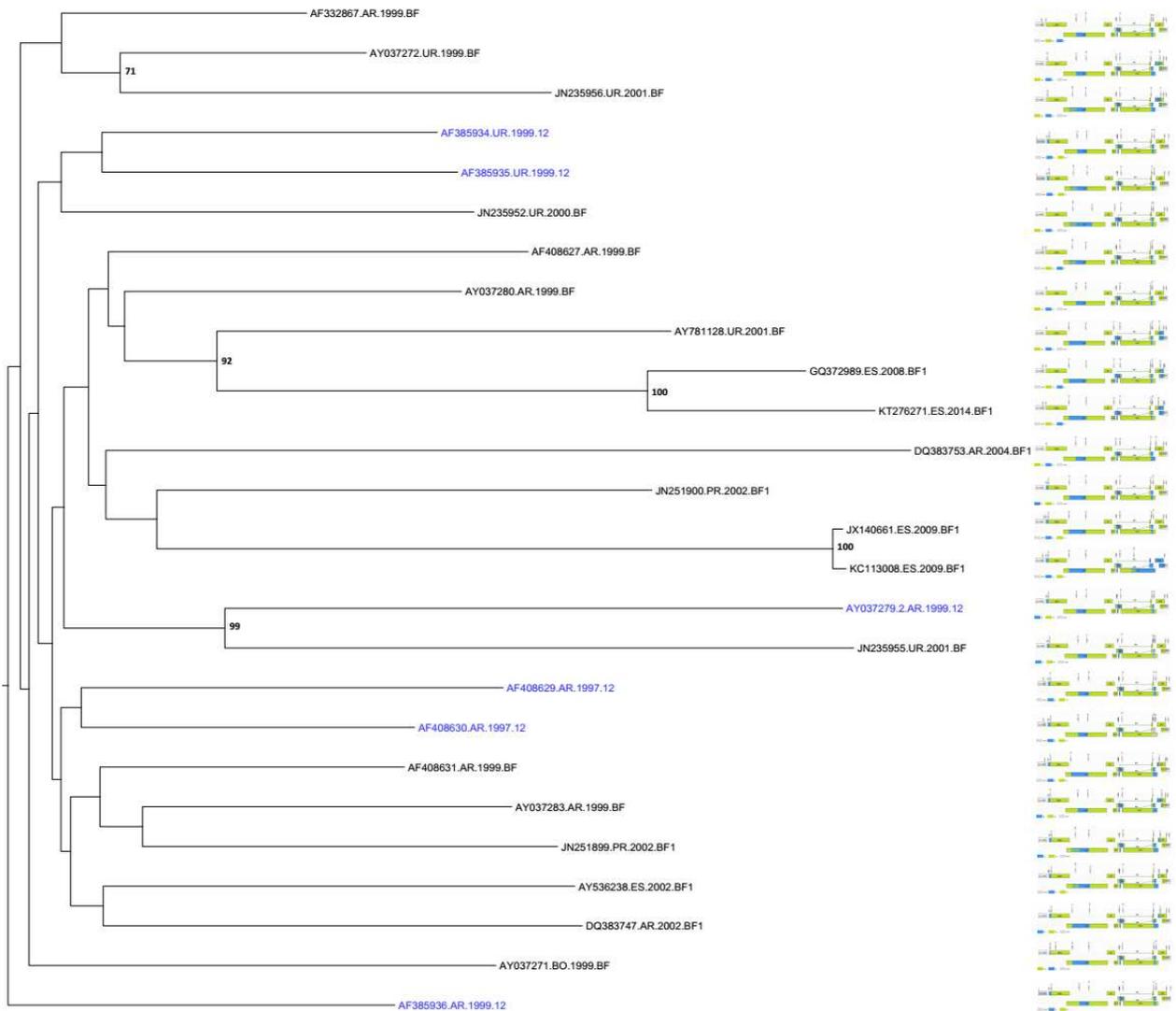


Fonte: Dados coletados pela autora.

Na **Figura 14** pode-se observar as seqüências classificadas como **CRF12** se apresentaram difusas na árvore juntamente com outras 20 seqüências relacionadas a esta forma recombinantes (AF332867.AR.1999.BF, AY037272.UR.1999.BF, JN235956.UR.2001.BF, JN235952.UR.2000.BF, AF408627.AR.1999.BF, AY037280.AR.1999.BF, AY781128.UR.2001.BF, GQ372989.ES.2008.BF1, KT276271.ES.2014.BF1, DQ383753.AR.2004.BF1, JN251900.PR.2002.BF1, JX140661.ES.2009.BF1, KC113008.ES.2009.BF1, JN235955.UR.2001.BF, AF408631.AR.1999.BF, AY037283.AR.1999.BF, JN251899.PR.2002.BF1, AY536238.ES.2002.BF1, DQ383747.AR.2002.BF1 e AY037271.BO.1999.BF).

Já foi demonstrado que a seqüência AY536238.ES.2002.BF1 tem o mesmo padrão de recombinação que a CRF12, portanto sugere-se a atualização desta seqüência no LA, onde a mesma encontra-se classificada como BF (SIERRA et al., 2005b)

**Figura 14** - Agrupamento da árvore ML retangular contendo a CRF12 e seqüências relacionadas.



Fonte: Dados coletados pela autora.

A análise filogenética destas seqüências mostrou-se robusta uma vez que as CRFs e as referências de cada subtipo formaram grupos relativamente homogêneos. Além disso, pode-se observar através da comparação com o padrão de recombinação destas seqüências que as URFs relacionadas agruparam juntas.

### 5.3 Análise de recombinação

Os padrões de recombinação das seqüências coletadas neste estudo, gerados pelo jpHMM, estão listados por ordem alfabética do número de acesso do Genbank (disponível no **Apêndice C**). Estes padrões foram também inseridos ao lado do nome das seqüências na árvore retangular ML (disponível no material suplementar em [goo.gl/nyKuKZ](http://goo.gl/nyKuKZ)).

Das 252 sequências avaliadas, 34,5 % (87/252) das sequências já foram inicialmente classificadas em alguma das CRF\_BF descritas e as demais como URFs. No entanto, algumas destas sequências apresentaram padrões de recombinação similares às CRFs já existentes, como discutido na análise da filogenia. Algumas sequências foram oriundas do mesmo paciente, sendo uma sequência caracterizada a partir do plasma e outra do genoma proviral (o par de sequências de cada paciente está separado por um travessão): AB253431.JP.ND.BF1/ AB253430.JP.ND.BF1, FJ670529.ES.2008.47/ KC473840.ES.2008.47, GQ372987.ES.2008.47/ KC473839.ES.2008.47, AB480298.JP.2004.BF1/ AB480299.JP.2004.BF1, AB480300.JP.2004.BF1/ AB480301.JP.2004.BF1 e JX140661.ES.2009.BF1/ KC113008.ES.2009.BF.

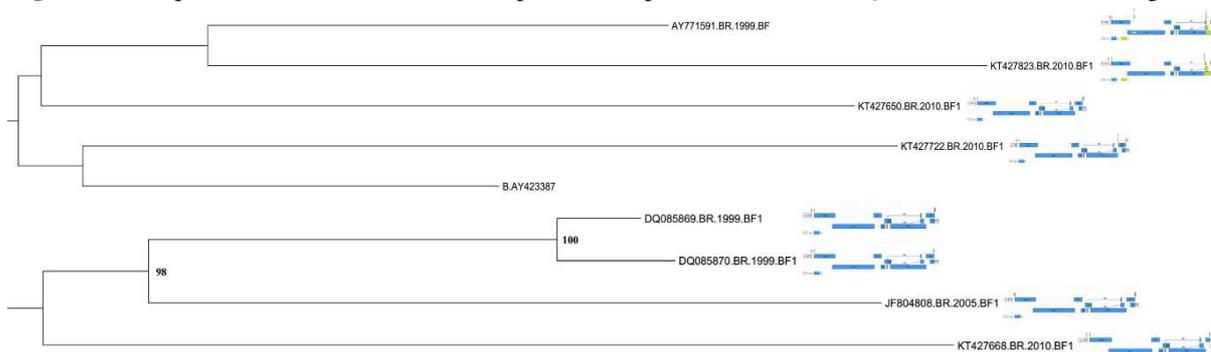
Já as sequências AY536233.CH.2001.BF e AY536234.CH.2001.BF foram provenientes de duas irmãs infectadas por via vertical e as sequências AY037275.AR.1999.BF e AY037277.AR.1999.BF de indivíduos que eram parceiros sexuais. Estas sequências possuem um padrão de recombinação idêntico ou muito próximo, o que era já era esperado devido a relação epidemiológica existente entre estes últimos dois pares de sequências citados.

No total, foram identificados 114 padrões de recombinação distintos entre os subtipos B e F nas sequências analisadas neste trabalho. Quatorze deles são referentes às CRFs já descritas e as sequências que apresentaram o mesmo padrão que estas CRFs, um da URF\_BF BF1G, quatro das URFs\_BF BCF e os demais referentes a URFs com uma ou duas sequências representantes. Esta variedade de padrões de recombinação indica que estes eventos de recombinação entre os subtipos B e F são frequentes e acarretam no surgimento de cepas virais viáveis e transmissíveis, ampliando ainda mais a diversidade genética do HIV-1.

### 5.3.1 Reclassificação de sequências previamente classificadas como BF

As sequências DQ085869.BR.1999.BF1, DQ085870.BR.1999.BF1, JF804808.BR.2005.BF1, KT427668.BR.2010.BF1, KT427650.BR.2010.BF1, KT427722.BR.2010.BF1 não foram identificadas no jpHMM como recombinantes BF e sim como subtipo B. Na **Figura 15** podem ser observados dois recortes da árvore que mostram os grupos formados por estas sequências, evidenciando inclusive, que uma sequência referência do subtipo B está no primeiro grupo.

**Figura 15** - Sequências classificadas como BF porém com padrão de recombinação B na árvore ML retangular.



Fonte: Dados coletados pela autora.

As sequências DQ085869.BR.1999.BF1 e DQ085870.BR.1999.BF1 foram descritas por De Sá Filho e colaboradores em 2007 como recombinantes B com mesmo padrão de recombinação: maior parte do genoma pertencente ao subtipo B com um fragmento na protease pertencente ao grupo F. No entanto o tamanho e a posição da inserção do fragmento F não foram informados. Em 2010, Sanabani e colaboradores demonstraram em uma publicação que esta classificação estava equivocada porém, até o momento, estas sequências continuam classificadas como recombinantes BF tanto no NCBI quanto em LA (SANABANI et al., 2010).

A sequência JF804808.BR.2005.BF1 foi descrita por Melo e colaboradores em 2012 como uma URF\_BF majoritariamente B com uma inserção do subtipo F na região *pol*. Porém, devido à baixa qualidade da imagem do padrão de recombinação publicado não foi possível identificar exatamente a posição da inserção. Interessantemente os autores também utilizaram o jpHMM para determinar o padrão de recombinação. Os autores não apresentaram análises filogenéticas que pudessem confirmar esta classificação (MELO; JAMAL; ZANOTTO, 2012). Na plataforma *Nucleotide* do NCBI a classificação desta sequência não é informada. Porém em LA ela também está classificada como BF1. Pode-se então concluir que, possivelmente, melhorias no algoritmo jpHMM levaram a correta classificação desta sequência na análise feita no atual estudo e que esta classificação é confirmada uma vez que esta sequência agrupou com outras sequências do subtipo B, que também foram classificadas equivocadamente, na árvore filogenética.

Estas três sequências citadas acima formam um grupo monofilético com um alto suporte estatístico (*bootstrap* de 98). Outra sequência, também relacionada com este grupo foi a sequência KT427668.BR.2010.BF. Estas sequência, juntamente com KT427722.BR.2010.BF1 foram publicadas em 2016 por Pêsoa e colaboradores. No LA estas sequências estão classificadas como recombinante BF1. No entanto, neste artigo o foco dos

autores foi descrever os recombinantes BF encontrados no trabalho e as referidas sequências não foram classificadas como recombinantes BF. Além disso, estas sequências estão classificadas como pertencentes ao subtipo B no NCBI, o que evidencia que foram equivocadamente classificadas no LA. Neste mesmo estudo, foi descrita uma outra sequência que apresentou divergência na classificação. Porém, diferentemente das sequências anteriores, KT427650.BR.2010.BF1 foi classificada como recombinante BF no estudo. Neste, o padrão de recombinação encontrado para esta sequência (utilizando o SIMPLOT e o jpHMM) foi um genoma majoritariamente B com duas inserções do subtipo F, uma na região *gag* e outra no início da região *pol*. No entanto, a referida sequência está classificada como subtipo B no LA (PESSÔA et al., 2016).

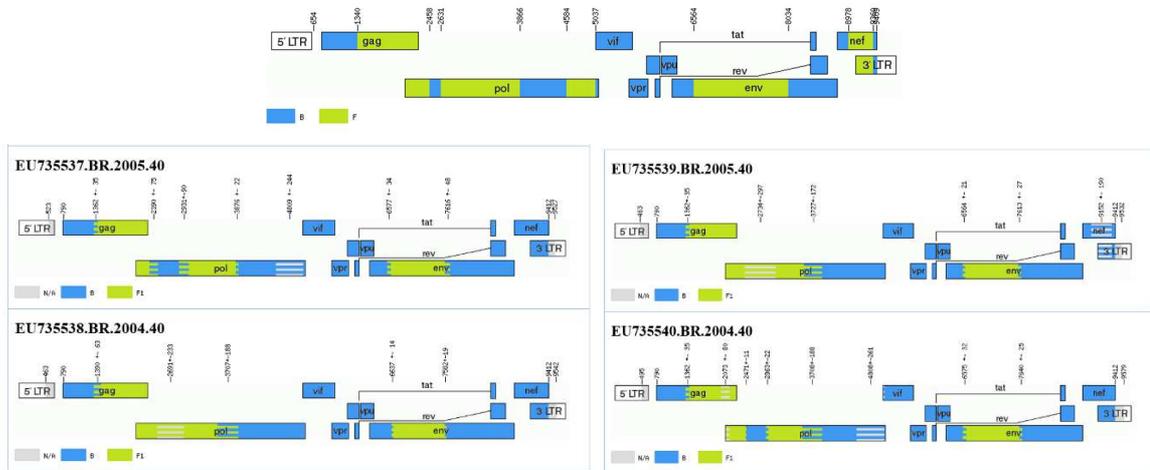
Portanto sugere-se que as sequências DQ085869, DQ085870 e JF804808 sejam reclassificadas como pertencentes ao subtipo B tanto em LA quanto no NCBI. Estas incoerências evidenciam a necessidade de uma curadoria mais cuidadosa para a submissão da sequência nos bancos de dados.

### 5.3.2 Discrepâncias na identificação dos padrões de recombinação

Os padrões de recombinação identificados pelo jpHMM apresentaram discrepâncias em algumas sequências. Em alguns casos, como o da sequência AY455780.BR.1999.BF, alguns fragmentos não são subtipados pelo algoritmo, sendo então descritos como N/A (“not available”). Esta é uma limitação reconhecida pelo próprio desenvolvedor para as análises realizadas na plataforma online. Em algumas sequências a não identificação do subtipo de alguns fragmentos pode ter sido causada pela presença de bases degeneradas na região, como o caso da sequência GU595154.IT.2002.BF. Provavelmente por causa deste mesmo motivo também ocorreu a identificação incorreta de determinados fragmentos, como por exemplo a sequência EU170152.LX.2003.42, e outras sequências da CRF42, que tiveram um fragmento identificado como subtipo A no final da sequência porém este fragmento, segundo os pontos de recombinação originais da CRF42, seria pertencente ao subtipo B.

Em outros casos, os padrões de recombinação divergiram um pouco dos pontos de recombinação previamente publicados, inclusive sequências classificadas na mesma CRF possuem padrão de recombinação distintos. Um exemplo deste caso acontece com a CRF40, como pode ser observado na **Figura 16** não qual representa o padrão de recombinação descrito para a CRF40 juntamente com os padrões identificados pelo jpHMM para as sequências classificadas com CRF40.

**Figura 16** - Padrão de recombinação oficial da CRF40 e o padrão encontrado individualmente para cada sequência que compõe a CRF40.



Fonte: Padrões individuais: gerados pelo jpHMM, dados do autor. O padrão oficial da CRF40 está disponível em <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html#CRF40>.

### 5.3.3 Investigação de pontos preferenciais (*Hotspots*) para recombinação no genoma do HIV-1

No intuito de investigar a existência de regiões genômicas preferenciais (*hotspots*) para a ocorrência de recombinação no HIV-1, os genes estruturais, acessórios e regulatórios foram analisados individualmente quanto à ocorrência de recombinação intragênica. O **Gráfico 2** mostra quantas sequências apresentaram recombinação intersubtipos ou não em cada gene e, das que não apresentaram recombinação, qual subtipo parental foi preservado naquele fragmento.

Pode-se observar que a região na qual a recombinação é mais frequente é a região *pol* (81%; 205/252), seguido das regiões *gag* (68%; 172/252) e *env* (54%/ 137/252). O número de sequências que possuem subtipos puros B ou F nestas regiões é similar. No entanto, as regiões *vpu*, *nef*, *rev*, *ltr*, *vpr* e *tat* são compostas, em sua maioria, pelo subtipo B. Portanto, há uma possibilidade do subtipo B ter vantagens adaptativas nestas regiões que justifiquem a manutenção do subtipo parental puro. Tem sido descrito na literatura que as razões para a alta taxa de recombinação em *pol* são a sua composição com bases homopoliméricas, similaridade da sequência e a presença de estruturas secundárias que favorecem a troca de fitas pela transcriptase reversa durante a geração do cDNA (SALMINEN et al., 1995; JETZT et al., 2000).

De Candia e colaboradores publicaram um trabalho mostrando que o *vpu* recombinante BF estruturalmente reorganizado tem, *in vitro*, a capacidade aumentada de aprimorar a replicação viral quando comparado à seu parental B (DE CANDIA et al., 2010).

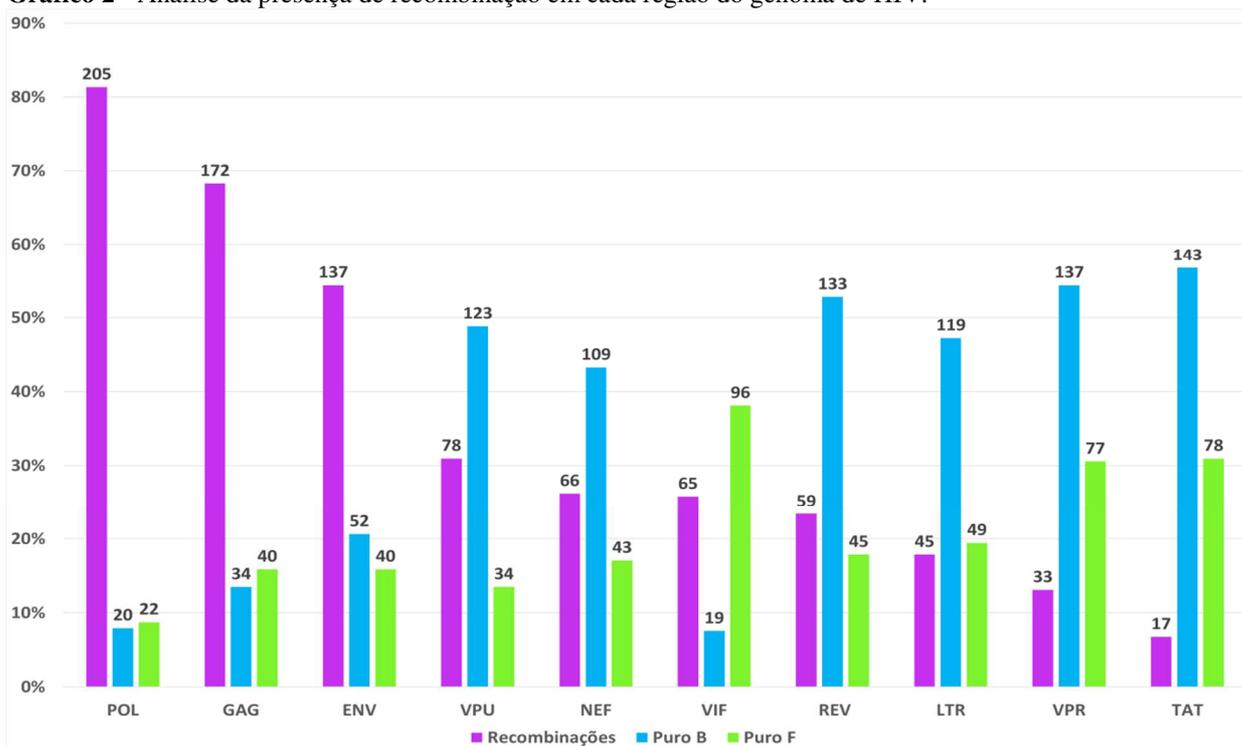
No entanto, como observado no gráfico abaixo, há uma maior quantidade de recombinantes BF com a região *vpu* com o subtipo B puro em detrimento da recombinação neste mesmo gene.

Interessantemente, apenas a região *vif*, um importante fator de infectividade viral que impede a ação de proteínas celulares da família APOBEC-3G (que tem a capacidade de tornar o HIV-1 não infeccioso através da indução de mutações G para A durante a síntese do DNA viral), na maioria das sequências (38%; 96/252), manteve preferencialmente o subtipo F puro. Seguindo o oposto da tendência das outras regiões, um menor número de sequências (8%, 19/252) mantem o subtipo B puro na região *vif*.

A maioria dos alelos de *vif* pertencentes ao subtipo F foram altamente efetivos em neutralizar a proteína A3H hapII enquanto os alelos de outros subtipos não B apresentaram baixa atividade para a mesma proteína (BINKA et al., 2012). Este achado também foi observado em outro estudo mais recente, mostrando que proteínas *vif* naturais do subtipos F1 atacam eficientemente as proteínas hA3G e hA3H hapII, enquanto são incapazes de degradar hA3C e hA3F (ZHANG et al., 2016). Pode-se inferir que a manutenção do subtipo F na região *vif* pode conferir uma maior vantagem a estes recombinantes. No entanto, ensaios funcionais de vírus contendo a região *vif* dos seus parentais preservadas e recombinantes BF poderão esclarecer a real importância da manutenção do subtipo F nesta região em vírus recombinantes BF.

Já nas regiões *tat* e *vpr* foram observados os menores números de recombinações (7%; 17/252 e 13%;33/252, respectivamente). Nestas regiões, a maioria das sequências mantem o fragmento do subtipo B puro (*tat* 57%; 143/252 e *vpr* 54%;37/252) seguido do F. Portanto, nestas regiões a recombinação é preterida em relação aos subtipos puros. Estes genes codificam proteínas que exercem funções essenciais para o ciclo viral como regulação da expressão gênica, transativação de genes celulares e indução da diferenciação celular (<https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>).

A análise do gene *env* mostrou que 21% (52/252) e 16% (40/252) das sequências possuíam o subtipo puro B e subtipo puro F nesta região, respectivamente. Aproximadamente metade das sequências (54%;137/252) apresentaram algum ponto de recombinação na região *env*. Buscando correlacionar a presença ou não de recombinação na região *env* e a utilização de determinado coreceptor na entrada viral foi observado que a maioria das sequências foram preditas como R5, independente do subtipo do fragmento *env* (*env* B: 59,5%, *env* F: 68,2% e *env* BF: 69,7%).

**Gráfico 2** - Análise da presença de recombinação em cada região do genoma de HIV.

Fonte: Dados coletados pela autora.

A região *nef*-U3, como já sugerido por (SANABANI et al., 2010) parece ser um importante ponto de recombinação (*hotspot*) pois a recombinação ocorre frequentemente nesta região, mais especificamente em torno da posição 9356 na CRF46 (n = 6), na CRF47 (n= 5) e ainda em outras 17 seqüências classificadas atualmente como URFs\_BF: KT427782.BR.2010.BF1, KT427774.BR.2010.BF1, KT427710.BR.2010.BF1, KT427705.BR.2010.BF1, JN235956.UR.2001.BF, HM026455.BR.2007.BF1, AY771591.BR.1999.BF, AY455784.BR.1994.BF, AY455783.BR.1999.BF, AY455782.BR.1999.BF1, AY455781.BR.1994.BF1, AF408626.AR.1999.BF, AF005495.BR.1993.BF1, AB480301.JP.2004.BF1, AB480300.JP.2004.BF1, AB480299.JP.2004.BF1 e AB480298.JP.2004.BF1, totalizando 28 (11%; 28/252) seqüências com recombinação BF nesta região.

Outra região na qual ocorre recombinação com frequência entre os subtipos B e F também foi identificada. Esta região é localizada no gene acessório *vif*, mais especificamente na região 5360 - 5390 em relação a cepa de referência HXB2. A CRF70 (n= 4) e a CRF71 (n=12) apresentam recombinação nesta região. Além destas, outras 19 seqüências classificadas como URFs\_BF, totalizando 35 (13,9%; 35/252) seqüências que apresentam ponto de recombinação nesta região: KJ849772.BR.2010.BF1, KT427758.BR.2010.BF1, AY771590.BR.2002.BF, KT427809.BR.2010.BF1, AY771592.BR.1999.BF,

KJ849757.BR.2010.BF1, GU595162 .IT.2001.BF1, GU595161.IT.2002.BF1,  
 DQ358799.BR.2001.BF1, AY771589.BR.1999.BF1, DQ358811.BR.2002.BF1,  
 GU595148.IT.2002.BF1, GU595156.IT.2001.BF1, KT427871.BR.2010.BF1,  
 KJ849797.BR.2010.BF1, AY771588.BR.1999.BF1, KT427672.BR.2010.BF1,  
 KT276259.ES.2014.BF1 e JN692456.BR.2004.BF1.

Portanto, foram identificados dois *hotspots*, um na região 5360 – 5390 do gene *vif* e outro em torno da posição 9356 do gene *nef*.

#### 5.4 Assinatura de nucleotídeos

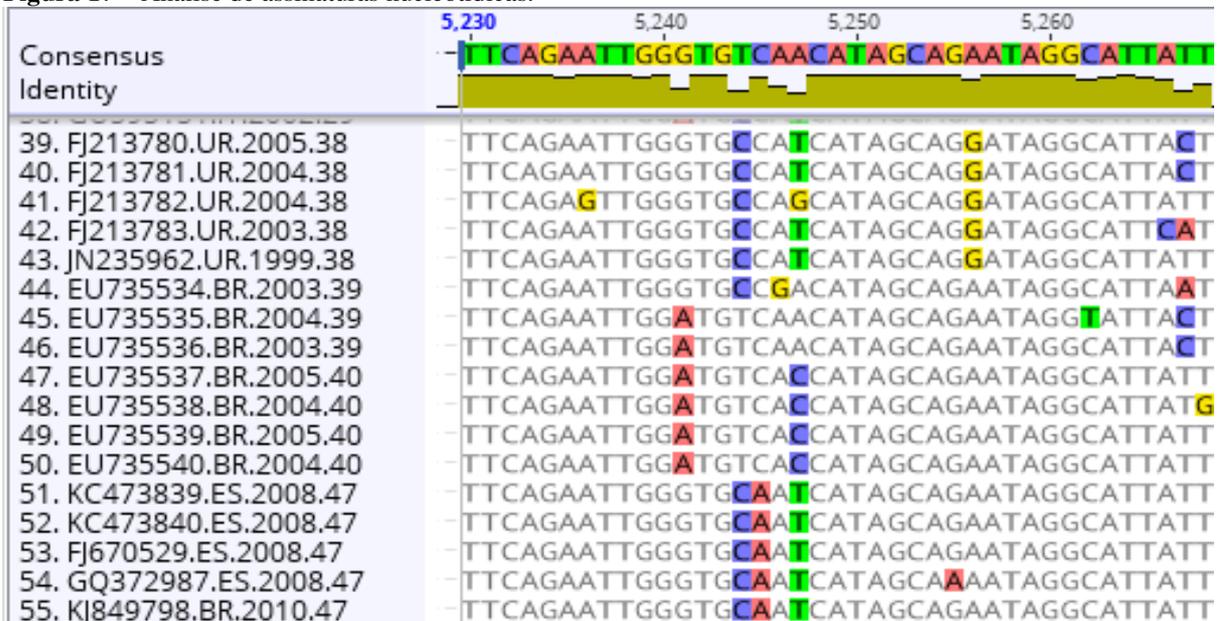
A busca por assinaturas nucleotídicas utilizando o programa Geneious retornou um total de 3628 polimorfismos em todo o alinhamento. Dentre estes, foram avaliados assinaturas nucleotídicas conservadas em todas as sequências de cada CRF e, posteriormente, as assinaturas nucleotídicas conservadas exclusivas de cada CRF. As assinaturas nucleotídicas conservadas descritas a seguir são exclusivas da respectiva CRF.

Foram identificados um total de 1391 posições com assinaturas nucleotídicas conservadas nas CRFs BF. Destas, 9 posições contendo assinaturas nucleotídicas conservadas em 7 CRFs\_BF, 27 posições em 6 CRFs\_BF, 83 posições em 5 CRFs\_BF, 127 em 4 CRFs\_BF, 187 em 3 CRFs\_BF, 285 em 2 CRFs\_BF e 673 posições com assinaturas nucleotídicas conservadas em apenas uma CRF, ou seja, são assinaturas nucleotídicas exclusivas da respectiva CRF. A CRF12 possui 10 assinaturas nucleotídicas conservadas exclusivas, a CRF17 possui 4, a CRF28 possui 27, a CRF29 possui 11, a CRF38 possui 15, a CRF39 possui 81, a CRF40 possui 74, a CRF42 possui 99, a CRF44 possui 129, a CRF46 possui 17, a CRF47 possui 177, a CRF70 possui 13, a CRF71 possui 3 e a CRF72 13. Estas assinaturas nucleotídicas conservadas exclusivas estão descritas no **Apêndice D**, identificadas com a CRF na qual foram identificadas, o número de sequências em cada CRF, o número de posições exclusivas conservadas encontradas na CRF, as posições nucleotídicas em relação a HXB2 e a respectiva base conservada.

A **Figura 17** exemplifica a análise de assinaturas nucleotídicas. Pode-se observar nesta imagem as sequências da CRF38, CRF39, CRF40 e CRF47. Na posição 5241 todas as sequências da CRF40 tem adeninas conservadas que foram identificadas como assinaturas nucleotídicas. Neste caso a posição e a base conservada foram anotadas. Nesta mesma posição pode-se observar que duas das três sequências da CRF39 também possuem uma adenina conservada identificada como assinatura nucleotídica. No entanto, como o critério de análise

foi a presença da assinatura nucleotídica conservada em todas as sequências da CRF esta posição não foi considerada para a CRF39. Já na posição 5223 houve a identificação de assinaturas nucleotídicas conservadas, citosinas mais especificamente, tanto na CRF38 quanto na CRF47. Nesta mesma posição pode-se observar ainda que a CRF40 tem timinas conservadas (coincidindo com a base consenso nesta posição) que não foram identificadas como assinaturas nucleotídicas. A posição 5256 mostra que as sequências da CRF38 tem G conservados que foram identificados como assinatura nucleotídica. Especificamente nesta imagem analisada, esta posição 5256G seria classificada com uma assinatura nucleotídica conservada exclusiva da CRF38, assim como a posição 5241A para a CRF40 e a posição 5245A para a CRF47.

**Figura 17** – Análise de assinaturas nucleotídicas.



Fonte: Dados coletados pela autora. Captura de tela da análise de assinatura molecular realizada no Genious. Posição 5230-5268 do alinhamento.

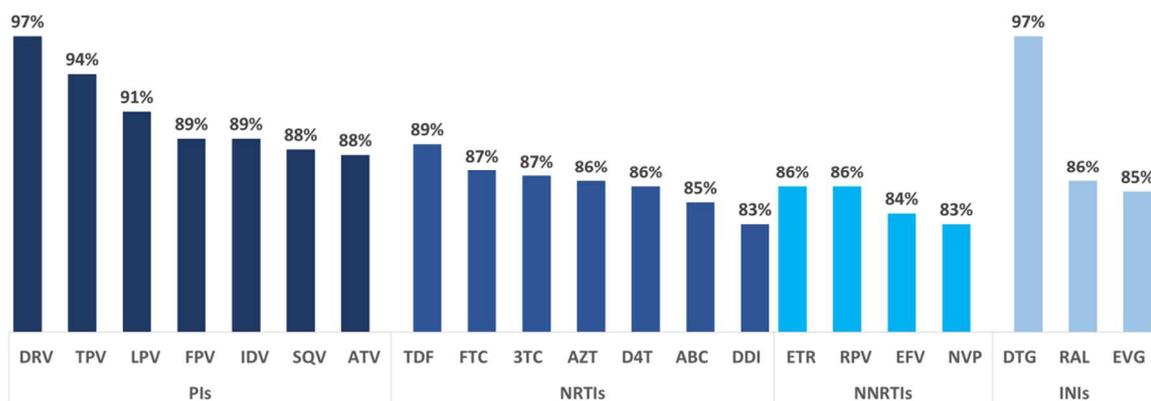
Apesar da grande diversidade de padrões de recombinação entre os subtipos B e F identificados neste estudo, muitos destes padrões oriundos de eventos distintos de recombinação, são extremamente semelhantes e difíceis de serem diferenciados pela análise filogenética. Sendo assim, a identificação de assinaturas nucleotídicas conservadas identificadas neste estudo nas diferentes CRFs poderá auxiliar na classificação de cepas recombinantes de BF HIV-1 que apresentam padrões repetitivos, bem como, na confirmação de pontos de recombinação.

### 5.5 Análise de resistência aos antirretrovirais

Nestas análises foi observado uma maior susceptibilidade das sequências estudadas aos PIs DRV (97%) e TPV (94%), ao NRTI TDF (89%), aos NNRTIs ETR e RPV e ao INI DTG (97%) (**Gráfico 3**). Já os menores índices de susceptibilidade observados para cada classe foram: 88% para PIs (ATV e SQV), 83% para NRTIs (DDI) e NNRTIs (NVP) e 85% para INIs (EVG).

Quanto a presença de mutações, um total de 43,7% (110/252) das sequências analisadas apresentaram alguma mutação que confere resistência a pelo menos uma das classes de ARVs analisadas, independentemente do nível de resistência conferido.

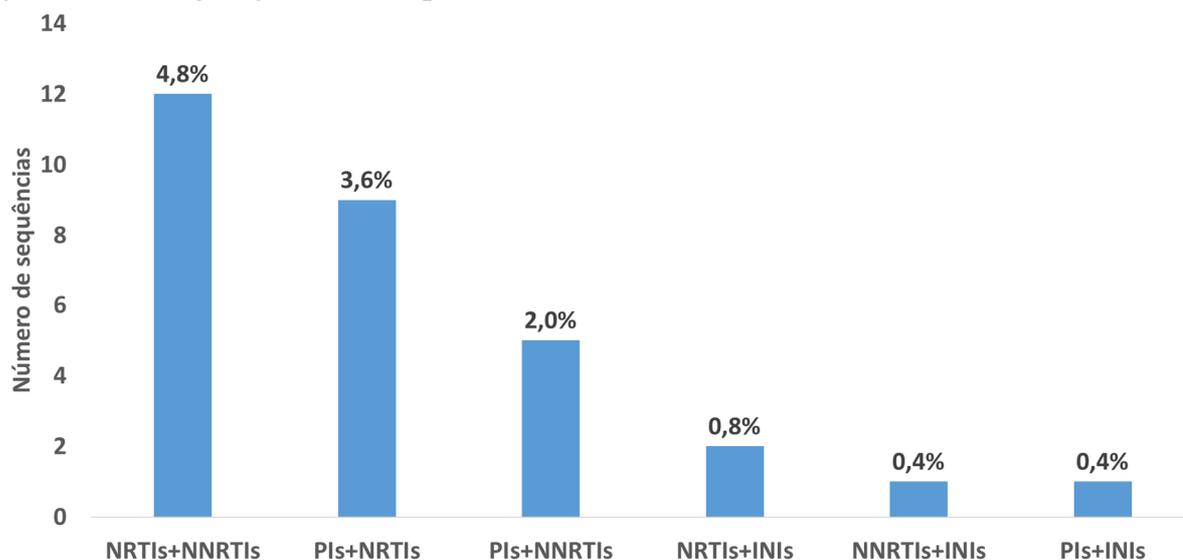
**Gráfico 3** – Porcentagem de susceptibilidade encontrada para as diferentes classes de ARVs estudados.



Fonte: Dados coletados pela autora.

Destas 110 sequências que apresentaram mutações, 64 (25,4%; 64/252) apresentaram mutações que conferem resistência a apenas uma classe de drogas independentemente do nível de resistência: 27 aos INIs (10,7%; 27/252), 20 aos NNRTIs (7,9%; 20/252), 10 aos PIs (4%; 10/252) e 7 aos NRTIs (2,8%; 7/252). Outras 31 sequências (12,3%; 31/110) apresentaram mutações a duas classes de ARVs, sendo a combinação NRTIs+NNRTIs a mais frequente (4,8%; 12/252), seguida pela combinação PIs+NRTIs (3,6%; 9/252), PIs+NNRTIs (2,4%; 6/252), NRTIs+INIs com 0,8% (2/252) e PIs+INIs ou NNRTIs+INIs com 0,4% cada (1/252), como mostrado no **Gráfico 4**.

**Gráfico 4** – Resistência a duas classes de ARVs. A porcentagem indica a frequência de sequências que apresentaram mutações que conferem algum nível de resistência à duas classes combinadas.



Fonte: Dados coletados pela autora.

Já 4,8% (12/252) apresentaram mutações à três classes de ARVs e 1,2% (3/252) às quatro classes analisadas. Encontrar mutações em quatro classes diferentes de drogas é algo incomum. As referidas sequências que as possuem são: GU595148.IT.2002.BF1, paciente do sexo masculino, bissexual, diagnosticado em 1990; HM026459.BR.2007.46, paciente do sexo feminino, 38 anos no momento da coleta, contagem de CD4 de 208 células/mm<sup>2</sup> e carga viral de 49240 cópias/mL; e JN251904.PR.2003.BF, paciente do sexo masculino, com 39 anos no momento da coleta.

### 5.5.1 Resistência aos PIs

Das 252 sequências 15,08% (38/252) apresentaram mutações na região da protease que conferem algum grau de resistência aos PIs. Nestas sequências foram encontradas um total de 27 diferentes mutações, sendo 14 delas classificadas como mutações principais (5,6%; 14/252) e 13 (5,2%; 13/252) como mutações acessórias (**Tabela 2**).

As mutações principais mais frequentes na região da protease foram L90M, D30N, V82A, I54V e M46I. A mutação L90M é selecionada principalmente pelo SQV, NFV, IDV e LPV enquanto a D30N é uma mutação não polimórfica selecionada pelo NFV levando a um alto nível de resistência a este fármaco. A mutação V82A, encontrada em oito sequências, é selecionada principalmente pelo IDV e LPV reduzindo a susceptibilidade aos mesmos, além de desencadear uma resistência cruzada ao ATV e NFV. Já a mutação I54V é selecionada principalmente pelo IDV e LPV, porém reduz a susceptibilidade a todos os PIs exceto DRV.

M46I é uma mutação que aumenta a eficiência catalítica da protease e geralmente é encontrada em combinação com outras mutações principais (<https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/>).

**Tabela 2** – Mutações encontradas na região da protease que conferem algum nível de resistência aos PIs.

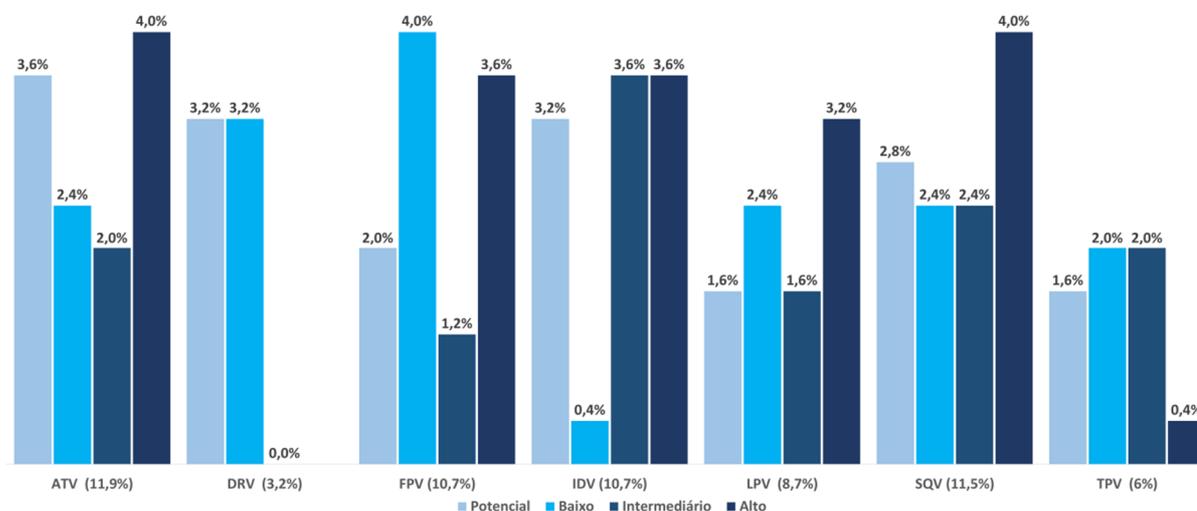
MUTAÇÕES NA PROTEASE					
Principais			Acessórias		
Substituição	N	%	Substituição	N	%
<b>L90M</b>	11	4,4%	<b>K20T</b>	9	3,6%
<b>D30N</b>	9	3,6%	<b>N88D</b>	9	3,6%
<b>V82A</b>	8	3,2%	<b>L33F</b>	4	1,6%
<b>I54V</b>	7	2,8%	<b>L10F</b>	3	1,2%
<b>M46I</b>	6	2,4%	<b>F53L</b>	3	1,2%
<b>I84V</b>	3	1,2%	<b>G48R</b>	2	0,8%
<b>L76V</b>	3	1,2%	<b>K43T</b>	2	0,8%
<b>M46MI</b>	3	1,2%	<b>G48E</b>	1	0,4%
<b>M46L</b>	2	0,8%	<b>G48GE</b>	1	0,4%
<b>I50L</b>	1	0,4%	<b>G48GR</b>	1	0,4%
<b>Q58E</b>	1	0,4%	<b>G73S</b>	1	0,4%
<b>V82F</b>	1	0,4%	<b>L24I</b>	1	0,4%
<b>V82IT</b>	1	0,4%	<b>Q58E</b>	1	0,4%
<b>D30DN</b>	1	0,4%			

Fonte: Dados coletados pela autora.

As mutações acessórias mais frequentes foram a K20T e N88D. Estas mutações estão associadas a diferentes níveis de resistência às drogas da classe dos inibidores de protease. A mutação K20T é uma mutação associada a uma susceptibilidade reduzida aos PIs, exceto ao DRV e TPV. Já a mutação N88D é selecionada pelo NFV e é usualmente encontrada com a mutação D30N, atuando aparentemente como uma mutação compensatória (<https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/>).

No **Gráfico 5** pode-se observar os níveis de resistência para cada droga inibidora da protease. Os índices de resistência mais elevados foram encontrados para o ATV (11,9%; 30/252) e SQV (11,5%; 29/252). No entanto, considerando apenas as mutações que conferem níveis de resistência intermediário ou alto (mais clinicamente significativas) o índice mais alto de resistência passa a ser atribuído ao IDV (7,1%; 18/252) seguido pelo SQV (6,3%; 16/252).

**Gráfico 5** – Níveis de resistência encontrados para os diferentes fármacos inibidores da protease.



Fonte. Dados coletados pela autora.

### 5.5.2 Resistência aos NRTIs e NNRTIs

Um total de 72 sequências (29,4%; 72/252) apresentou alguma mutação na região da transcriptase reversa (**Tabela 3**). Mutações que conferem algum nível de resistência aos NRTIs foram identificadas em 17,5% (44/252) das sequências, sendo as mutações M41L, T215Y, D67N, M184V, K70R e L210W as mais frequentes. M41L usualmente ocorre com a mutação T215Y e juntas conferem alto nível de resistência ao AZT e D4T e nível de resistência baixo ou intermediário ao ABC, DDI e TDF. A mutação M184V é uma mutação associada a uma replicação viral reduzida tanto *in vitro* quanto *in vivo*. Esta mutação é selecionada pelos fármacos 3TC e FTC e reduz a susceptibilidade a estas drogas em mais de 100 vezes, entretanto, está associada a um aumento da susceptibilidade ao AZT, D4T e TDF. K70R confere nível de resistência intermediário ao AZT e baixo nível de resistência aos D4T e TDF. Já a mutação L210W usualmente é encontrada em combinação com as mutações M41L e T215Y e juntas conferem alto nível de resistência ao AZT e D4T (<https://hivdb.stanford.edu/dr-summary/resistance-notes/NRTI/>).

As mutações para a classe dos NNRTIs foram identificadas em 52 (20,63%; 52/252) sequências e as mais frequentes foram K103N, E138A, G190A e V108I. A mutação K103N é selecionada pelos fármacos NVP e EFV reduzindo a susceptibilidade destes fármacos em torno de 50 e 20 vezes, respectivamente. E138A é uma mutação fracamente selecionada pelos fármacos ETR e RPV. Já a mutação G190A é selecionada pelo NVP e EFV, enquanto a

mutação V108I é selecionada *in vitro* pelo NVP e EFV e em pacientes tratados com NVP, EFV e ETR (<https://hivdb.stanford.edu/dr-summary/resistance-notes/NNRTI/>).

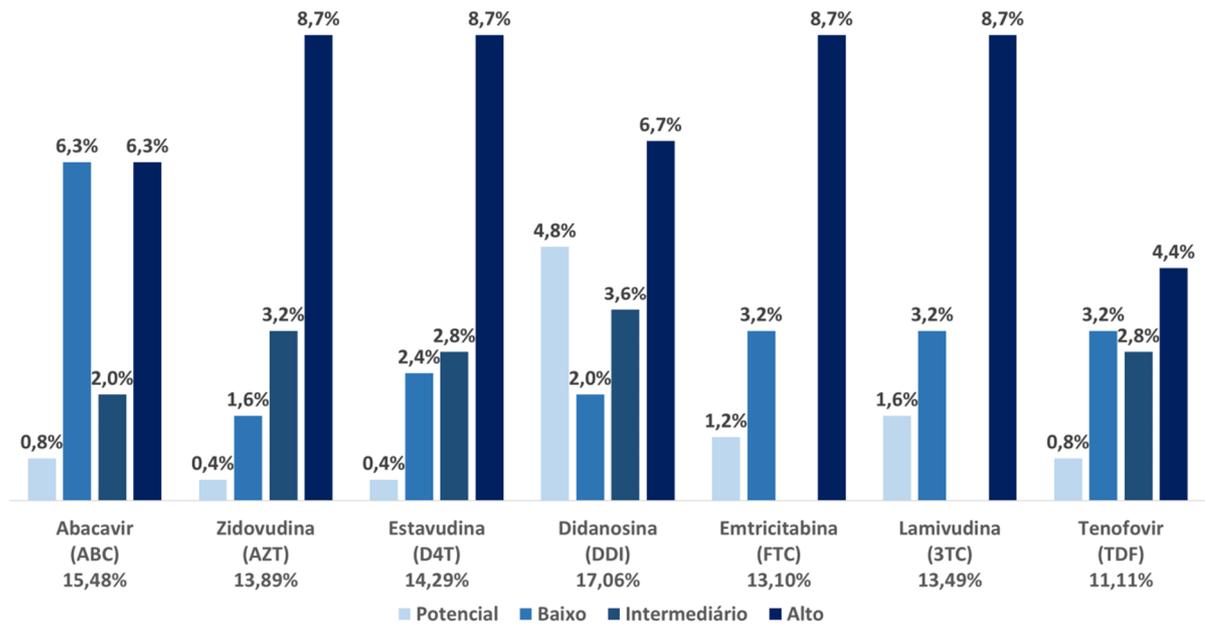
Observa-se que os níveis de resistência mais elevados para a classe dos NRTIs (**Gráfico 6**) foram encontrados para as drogas AZT, D4T, FTC e 3TC, com 8,7% (22/252, cada) das sequências apresentando mutações que conferem alto nível de resistência às referidas drogas. Destes fármacos, o TDF e 3TC são utilizados como medicamentos de primeira linha no tratamento no Brasil porém os níveis de resistência pra estes fármacos nas sequências avaliadas ainda são considerados baixos sugerindo que eles tendem a ser efetivos em indivíduos infectados com o subtipo BF (MINISTÉRIO DA SAÚDE, 2015).

**Tabela 3** – Mutações encontradas na região da transcriptase reversa (RT) que conferem algum nível de resistência aos NRTIs ou NNRTIs.

MUTAÇÕES NA RT					
NRTI			NNRTI		
Substituição	N	%	Substituição	N	%
M41L	20	7,9%	K103N	14	5,6%
T215Y	20	7,9%	E138A	8	3,2%
D67N	17	6,7%	G190A	6	2,4%
M184V	17	6,7%	V108I	6	2,4%
K70R	13	5,2%	N348I	5	2,0%
L210W	12	4,8%	Y181C	5	2,0%
K219Q	7	2,8%	A98G	3	1,2%
T215F	7	2,8%	H221Y	3	1,2%
E44D	6	2,4%	K238T	3	1,2%
L74V	3	1,2%	M230I	3	1,2%
M184I	3	1,2%	V179D	3	1,2%
D67G	2	0,8%	E138K	2	0,8%
E40F	2	0,8%	K103S	2	0,8%
E44A	2	0,8%	L100I	2	0,8%
K219E	2	0,8%	V106A	2	0,8%
L74I	2	0,8%	E138EK	1	0,4%
T69D	2	0,8%	F227L	1	0,4%
V75M	2	0,8%	G190GA	1	0,4%
F77L	1	0,4%	G190GEKR	1	0,4%
K219R	1	0,4%	G190R	1	0,4%
K70*	1	0,4%	K101E	1	0,4%
M184MI	1	0,4%	K101PQ	1	0,4%
M184MV	1	0,4%	K103KE	1	0,4%
T215D	1	0,4%	K103KI	1	0,4%
T215E	1	0,4%	K103KN	1	0,4%
T215S	1	0,4%	P236L	1	0,4%
T215TNSY	1	0,4%	V179E	1	0,4%
V75A	1	0,4%	V179T	1	0,4%
V75T	1	0,4%			

Fonte: Dados coletados pela autora. \*: Deleção

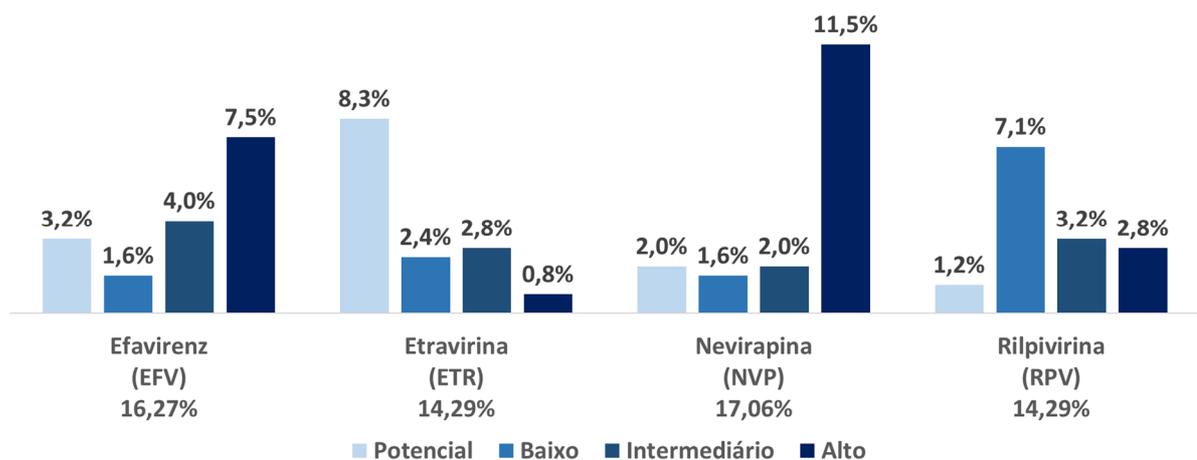
**Gráfico 6** – Níveis de resistência encontrados para os diferentes fármacos da classe dos NRTIs.



Fonte: Dados coletados pela autora.

O **Gráfico 7** mostra os níveis de resistência encontrados para a classe dos NNRTIs. Observa-se que 11,5% (29/252) das sequências apresentaram alto nível de resistência ao NVP e 7,5% (19/252) ao EFV. Este último, juntamente com o TDF e o 3TC compõem o tratamento de primeira linha aqui no Brasil.

**Gráfico 7** – Níveis de resistência encontrados para os diferentes fármacos da classe dos NNRTIs.



Fonte: Dados coletados pela autora.

### 5.5.3 Resistência aos INIs

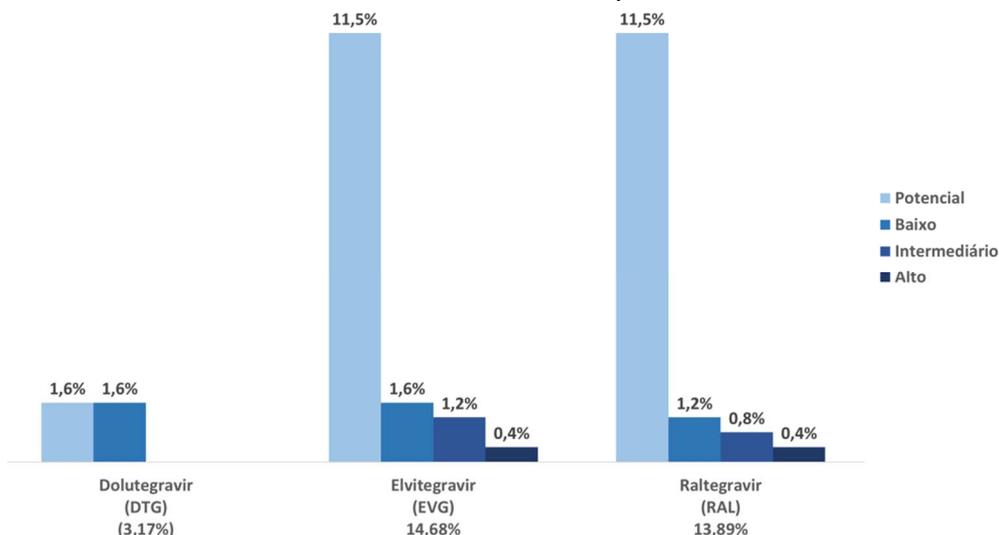
Os INIs são considerados drogas fundamentais para os esquemas de resgate, principalmente no tratamento de terceira linha, após falha das três classes PIs, NRTIs e NNRTIs, comumente utilizadas. Das sequências estudadas, 15,9% (40/252) apresentaram mutações na região da integrase. Das mutações encontradas nesta região, apenas 10 (4%; 10/252) são mutações principais e as demais são classificadas como mutações acessórias 41 (16,3%; 41/252) (**Tabela 4**). Destas, as mutações mais frequentes, G163R/K, já foram descritas como mutações com alta prevalência em recombinantes BF (DELICADO; TORRECILLA; HOLGUÍN, 2016).

**Tabela 4** – Mutações encontradas na região da integrase que conferem algum nível de resistência ao INIs.

MUTAÇÕES NA INTEGRASE					
Principais			Acessórias		
Substituição	N	%	Substituição	N	%
<b>E138EK</b>	2	0,8%	<b>G163R</b>	19	7,5%
<b>E92K</b>	2	0,8%	<b>G163K</b>	11	4,4%
<b>E138K</b>	1	0,4%	<b>R263RK</b>	2	0,8%
<b>N155D</b>	1	0,4%	<b>G140*</b>	1	0,4%
<b>N155*</b>	1	0,4%	<b>G140R</b>	1	0,4%
<b>P145PS</b>	1	0,4%	<b>G163EK</b>	1	0,4%
<b>T66TP</b>	1	0,4%	<b>G163KR</b>	1	0,4%
<b>Y143YH</b>	1	0,4%	<b>S153SF</b>	1	0,4%
			<b>S153Y</b>	1	0,4%
			<b>T97A</b>	1	0,4%
			<b>T97TA</b>	1	0,4%
			<b>A128T</b>	1	0,4%

Fonte: Dados coletados pela autora. \*: Deleção.

**Gráfico 8** – Níveis de resistência encontrados para os diferentes fármacos da classe dos INIs.



Fonte: Dados coletados pela autora.

Portanto, como já seria esperado, em virtude da maioria das mutações terem sido identificadas como acessórias, pode-se observar que estas mutações conferem, em sua maioria, uma resistência potencial ao EVG e ao RAL (11,51% cada; 29/252) (**Gráfico 8**). O elevado índice de sequências com mutações com nível de resistência potencial pode ser explicado exatamente pelo fato de a maioria das mutações encontradas serem mutações acessórias.

A escolha da classe e de qual fármaco da classe será utilizado no tratamento, bem como a combinação entre eles varia de acordo com as estratégias de saúde de cada país. Além disso as informações sobre o estado sintomático e tratamento dos pacientes dos quais estas sequências foram oriundas foram escassos. Apenas 8% (21/252) das sequências possuíam informação sobre a utilização ou não de tratamento ARV e destes 21 apenas 8 pacientes (3,2%; 8/252) estavam em tratamento. As informações disponíveis sobre o tratamento destes oito pacientes referiam-se apenas às classes de drogas utilizadas no tratamento.

Em virtude desses fatores há uma limitação nas associações realizadas nas análises de resistência aos ARVs. No entanto, embora estes dados não sejam representativos de uma população específica, os dados deste trabalho apresentam grande relevância uma vez que podem ser úteis no manejo clínico.

### **5.6 Predição do coreceptor de entrada viral**

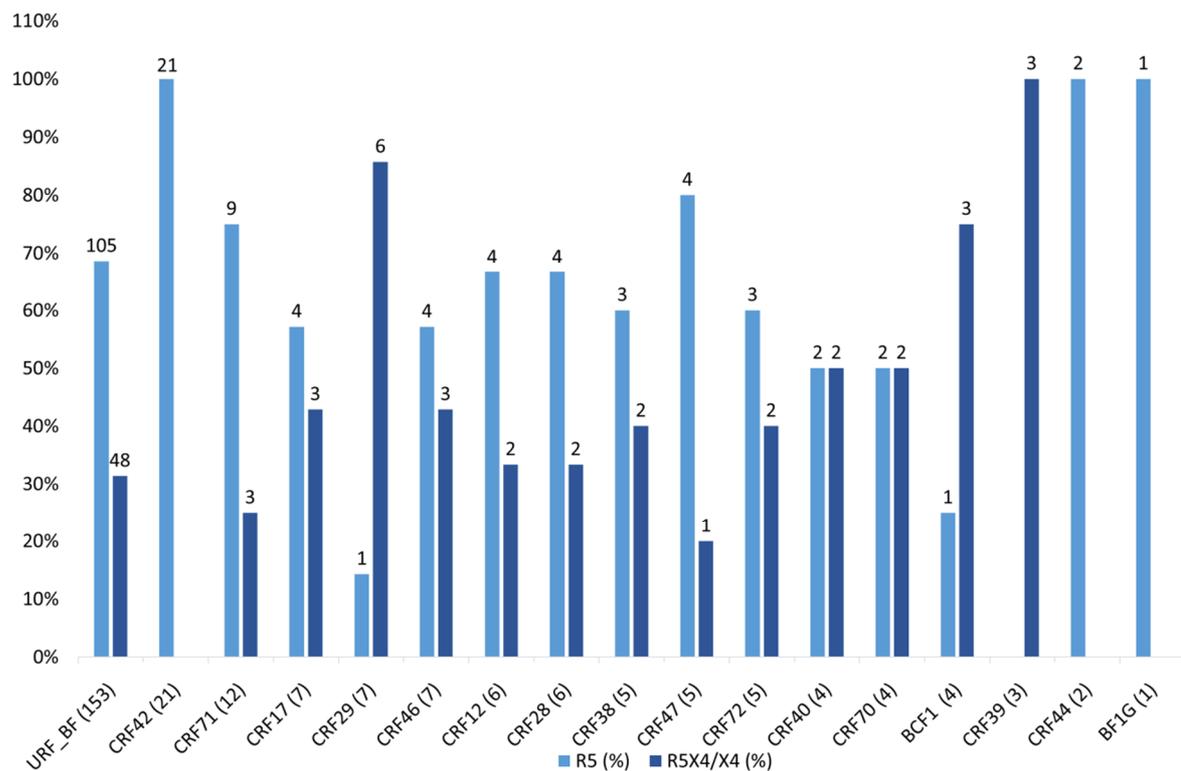
A alça V3 tem um importante papel nas interações com o receptor e o coreceptor que determinam o tropismo e a entrada viral. A predição do uso do coreceptor indicou que, provavelmente, 67,5% (170/252) das sequências foram provenientes de vírus que são capazes de utilizar o CCR5 como coreceptor no processo da entrada viral enquanto 32,5% (82/252) são classificados como X4 ou duplo trópicos (R5/X4) (**Gráfico 9**). Neste gráfico, pode-se observar a prevalência da utilização de cada coreceptor de acordo com a classificação das sequências.

Vírus capazes de utilizar o coreceptor CCR5 são mais comuns em infecções iniciais e tem uma maior taxa de transmissão do que vírus X4 enquanto os vírus X4 ou R5/X4 estão associados com uma rápida progressão da doença (CUEVAS et al., 2010). No entanto, não foi possível aprofundar esta comparação no presente estudo pois não há dados clínicos suficientes dos pacientes dos quais as sequências foram coletadas.

Pode-se observar no gráfico acima que as sequências da CRF39 foram todas classificadas com R5X4/X4 enquanto as sequências da CRF42 e a CRF44 foram classificadas

como R5. A análise estatística destes dados mostra que há diferença estatística quanto o uso do coreceptor utilizado na CRF42 ( $p = 0.0003$ ), CRF29 ( $p = 0.0054$ ) e na CRF39 ( $p = 0.0336$ ), mostrando que as sequências destas CRFs possuem uma preferência no uso do coreceptor CCR5 no processo da entrada viral, como já observado para as 252 sequências no total.

**Gráfico 9** – Prevalência da predição de vírus R5 ou R5X4/X4 de acordo com a classificação.

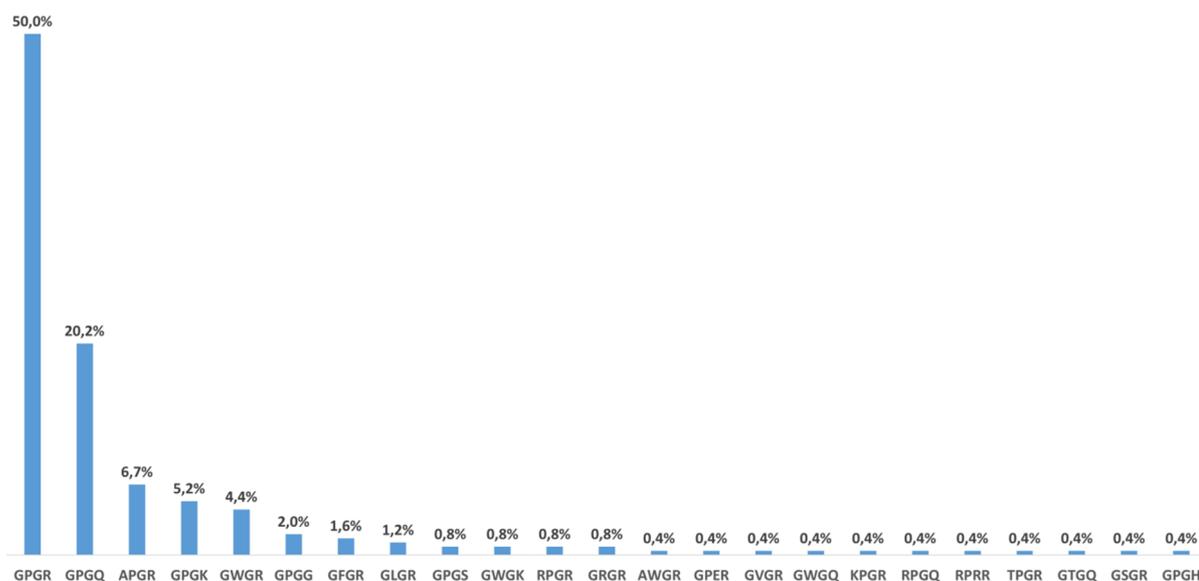


Fonte: Dados coletados pela autora. O número após a classificação da sequência indica o N total e os números acima das barras indicam N de cada predição.

Todos os subtipos podem usar ambos coreceptores porém tem sido demonstrado que vírus do subtipo D são em sua maioria duplo trópicos e que os subtipos C e G tem uma taxa menor de vírus X4 do que o subtipo B. Sabe-se também que os vírus que utilizam o coreceptor CXCR5 são mais transmissíveis o que poderia indicar uma vantagem aos recombinantes BF uma vez que a maioria destas sequências analisadas foram classificadas como R5 (67,5%) (CUEVAS et al., 2010). Além disso, considerando que 67,5% das sequências analisadas foram preditas como provenientes de vírus que utilizam exclusivamente o coreceptor CCR5, pode-se avaliar que provavelmente a grande maioria dos pacientes infectados com recombinantes BF responderiam ao Maraviroc, droga inibidora do coreceptor CCR5.

Na análise do loop da alça V3 foram encontrados 23 diferentes motivos estruturais do tetrapeptídeo da alça V3, sendo que em 3 sequências (1,2%; 3/252) não foi possível identificá-los por causa de bases degeneradas na sequência (**Gráfico 10**). Todos motivos estruturais encontrados já foram descritos na literatura.

**Gráfico 10** – Frequência de cada motivo estrutural do tetrapeptídeo da alça V3 da proteína gp120 encontrado nas sequências estudadas.

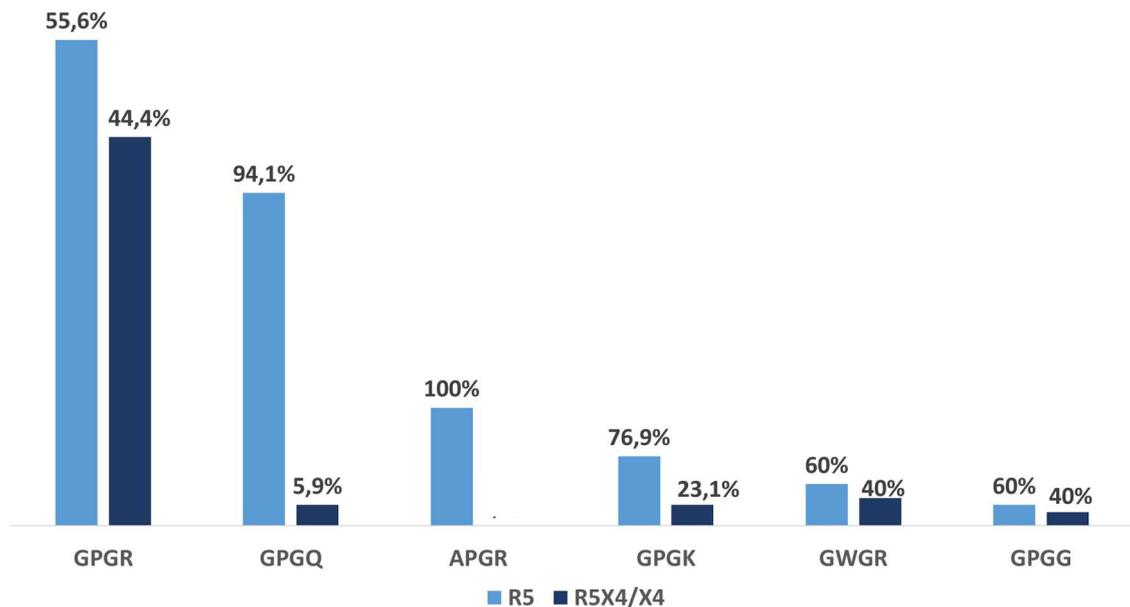


Fonte: Dados coletados pela autora.

A grande variedade de motivos estruturais mostra a complexidade e variabilidade nesta região. Dentre os tetrapeptídeos mais frequentes (GPGR 50%; GPGQ 20,2%; APGR 6,7%; GPGK 5,2%; GWGR 4,4%; e GPGG 2%), a maioria está relacionado com o uso preferencial do coreceptor CXCR5, como mostrado no **Gráfico 11**.

É importante salientar que o motivo GWGR é conhecido como um domínio característico do subtipo B brasileiros. Neste estudo, o mesmo foi encontrado em 10 sequências (2 CRF28, 2 CRF29 e em 6 URFs\_BF). As CRFs 28 e 29 foram identificadas no Brasil e parecem terem sido originadas de ancestrais brasileiros. As URFs nas quais este motivo foi encontrado possuem a região V3 pertencente ao subtipo B e são quatro sequências do Brasil (AY771590.BR.2002.BF, JN692448.BR.2003.BF1, KT427758.BR.2010.BF1, KT427828.BR.2010.BF1) e duas sequências da Argentina: AF408628.AR.1999.BF e AY037267.AR.2000.BF, como já descrito anteriormente (VILLANOVA, 2010).

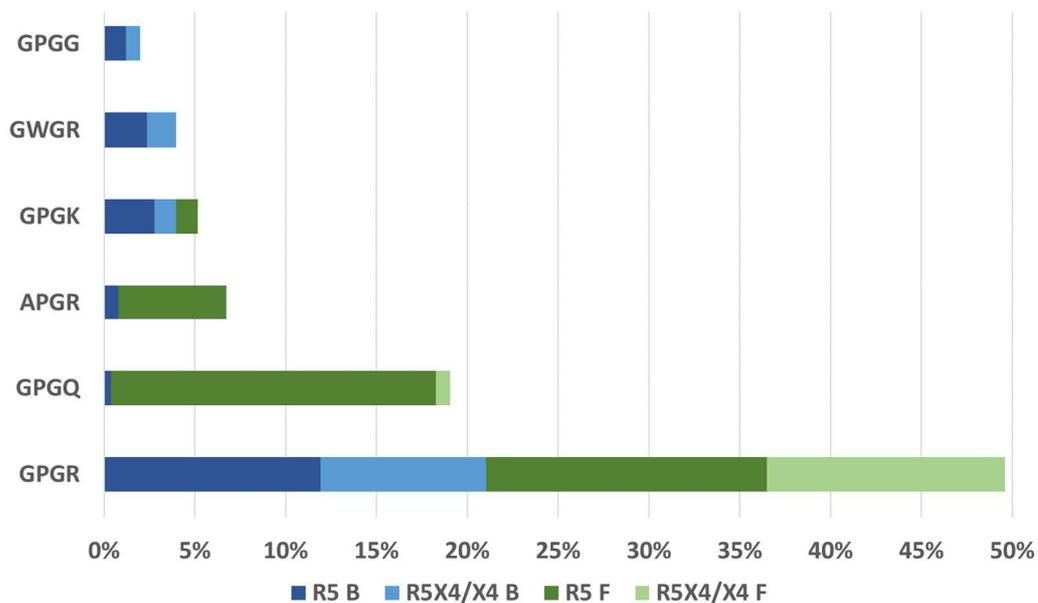
**Gráfico 11** – Frequência do uso de coreceptor predito de acordo com o motivo estrutural encontrado na alça V3 da proteína gp120.



Fonte: Dados coletados pela autora.

A análise do subtipo da região da alça V3 mostrou que o subtipo F é predominantemente encontrado nesta região (57,5%; 146/252) enquanto o subtipo B é encontrado em 39,7% (100/252) das sequências. Das sequências com subtipo F na alça V3, 74% (108/146) foram classificadas como R5 enquanto 60% (58/100) das classificadas como subtipo B na referida região foram classificadas como R5.

**Gráfico 12** - Classificação das sequências com os motivos mais frequentes na alça V3 de acordo como subtipo desta região e o coreceptor predito.



Fonte: Dados coletados pela autora.

Pode-se observar no **Gráfico 12** que as sequências que possuem o motivo GPGR, independentemente da predição do coreceptor de entrada viral, possuem número de sequência equivalentes do subtipo B e F na alça V3. No entanto, as sequências com os motivos GPGQ e APGR utilizam predominantemente o coreceptor CXCR5 e as alças V3 destas sequências são majoritariamente do subtipo F, (17,9%; 45/252 – GPGQ, R5, F<sup>V3</sup>) e (6%; 15/252 – APGR, R5, F<sup>V3</sup>).

O motivo GWGR foi anteriormente associado a utilização exclusiva do receptor CCR5 (LEAL et al., 2008b). Porém, outros trabalhos já relataram que vírus com este motivo estrutural na alça V3 podem utilizar o receptor CXCR4 (X4 ou R5X4) a uma taxa de aproximadamente 25% (FERRARO et al., 2001; VILLANOVA, 2010; SANABANI et al., 2011). No entanto, como pode ser observado no **gráfico 11**, cerca de 40% das sequências que apresentam o motivo estrutural GWGR na alça V3 foram preditas como R5/X4 evidenciando que nas sequências BF analisadas não há uma relação direta entre a presença deste motivo e a utilização exclusiva ou preferencial do coreceptor CXCR5.

## 5.7 Reclassificação de sequências

Durante o desenvolvimento deste estudo foi identificada a necessidade de reclassificação de algumas sequências e a revisão da classificação nos bancos de dados. Estas observações estão listadas brevemente abaixo e sugere-se a correção das mesmas.

A sequência KT427774 está classificada como F1 em LA porém sugere-se a correção para BF. Já a sequência KJ849791 foi reclassificada com BF1 por Pessoa porém continua classificada como F1 em LA (PESSÔA et al., 2015). As sequências DQ358802.BR.2001.46 e DQ358801.BR.2001.46, anteriormente classificadas como F1, foram reclassificadas como CRF46 em 2010 (SANABANI et al., 2010) porém continuam identificadas como subtipo F1 no NCBI.

Sugere-se também a reclassificação das sequências AY455781.BR.1994.BF1, HM026455.BR.2007.BF1, AY455782.BR.1999.BF1, KT427779.BR.2010.BF1, KT427774.BR.2010.BF1 AB480298.JP.2004.BF1, AB480299.JP.2004.BF1, AB480300.JP.2004.BF1 e AB480301.JP.2004.BF1 para CRF46, como já apontado por Sanabani para algumas destas sequências (SANABANI et al., 2010).

Estas sugestões de reclassificação serão encaminhadas para LA e o NCBI.

## 5.8 Considerações finais

A diversidade de padrões de recombinação em cepas de HIV-1 encontrados nesse estudo, retrata a complexidade da epidemia de HIV no Brasil e mostra que os eventos de recombinação estão ocorrendo com frequência e que estes recombinantes apresentam um importante papel na epidemia do país. Segundo Sierra e colaboradores, a alta prevalência de formas recombinantes deriva de dois mecanismos. O primeiro seria a partir da expansão das CRFs para proporções epidêmicas e a segundo mecanismo seria a continua geração de formas recombinantes únicas entre diferentes variantes, tanto entre subtipos quanto entre as formas recombinantes (SIERRA et al., 2005a).

Diversos fatores podem contribuir para a alta diversidade de recombinantes encontrados como, por exemplo, a origem destes recombinantes desde o início da epidemia do HIV. Estima-se que a origem dos recombinantes BF no Brasil ocorreu entre 1984 e 1991 (LEAL et al., 2008a; MICHAEL M. THOMSON et al., 2002). Como estes recombinantes estão em circulação há um longo período de tempo, há um aumento na chance de recombinação, especialmente no Brasil, onde tanto o subtipo B (desde 1983) quanto o F (desde 1989) também já cocirculam há bastante tempo e, portanto, há um aumento da probabilidade de recombinações de segunda geração ocorrerem. Como observado neste trabalho, os recombinantes BF tem crescido tanto em CRFs, com o maior número de CRFs identificadas, quanto em número e diversidade de URFs. Portanto pode-se atribuir a alta prevalência de recombinantes BF aos dois mecanismos citados.

Como já discutido por alguns autores, o aumento da prevalência de recombinantes BF tende a levar ao desaparecimento do seu parental F, que já é normalmente encontrado em baixa frequência (SANABANI et al., 2006). Uma observação secundária a esta é que há, provavelmente, uma alta taxa de coinfeção entre os subtipos B e F uma vez que há uma grande variedade de URFs<sub>BF</sub> e que, para os eventos de recombinação acontecerem, é necessário que o indivíduo esteja coinfectado.

A grande diversidade de perfis de recombinação encontrados entre os subtipos B e F aumenta a probabilidade destes recombinantes adquirirem novas propriedades biológicas que confirmam vantagens adaptativas em relação aos subtipos parentais, uma vez que que cepas mais adaptadas podem surgir a partir do processo de recombinação. E, deve-se ainda considerar, que o número de recombinantes BF identificados pode estar subestimado em função da classificação de sequências parciais serem baseadas apenas em sequenciamento parcial fragmentos pequenos do genoma.

Apesar de oito CRFs\_BF terem sido caracterizadas no Brasil, elas não representam a maioria das sequências encontradas. Ao contrário, há um aumento no número e diversidade de URFs identificados (DE SÁ FILHO et al., 2007; SANABANI et al., 2010; PESSOA et al., 2014a; PESSÔA et al., 2014b). Este aumento na diversidade genética viral tem repercussões potenciais no desenvolvimento de vacinas e incertezas sobre a resposta terapêutica, por exemplo. Este estudo contribui na descrição da diversidade genética de formas recombinantes BF e portanto poderá ser utilizado como fonte de embasamento para o desenvolvimento de vacinas para regiões onde os subtipos B e F cocirculam e/ou regiões na qual o subtipo BF tem um importante representação na epidemiologia local.

Portanto, evidencia-se a importância de monitorar a epidemiologia molecular da infecção pelo HIV para avaliar o impacto destes recombinantes na progressão e controle da epidemia do HIV.

## 6 CONCLUSÕES

Foram encontradas 252 sequências de 13 países diferentes que possuem, em sua maioria, poucas informações disponíveis sobre os pacientes dos quais os isolados virais foram coletados. As análises filogenéticas, juntamente com o padrão de recombinação destas sequências e a correlação com dados já disponíveis na literatura permitiu a identificação de algumas sequências que devem ser reclassificadas e outras que devem ter a classificação corrigida nos bancos de dados. Além disso, pode-se observar as relações entre as URFs entre si e com as CRFs, das CRFs entre si e de ambos com seus parentais.

A grande diversidade dos padrões de recombinação, representada pelos 114 padrões de recombinação distintos observados, mostra que a recombinação entre os subtipos B e F é frequente. Além disso, deve-se considerar ainda que a detecção de recombinantes BF pode estar subestimada uma vez que é comum a subtipagem baseada apenas em um fragmento genômico. Pode-se observar também que a região *vif* da maioria das sequências tem o subtipo F puro, em contraste com as regiões *vpu*, *nef*, *rev*, *ltr*, *vpr* e *tat* nas quais a maioria das sequências conservam o subtipo B. Já as regiões *pol*, *gag* e *env* da maioria das sequências apresentam pontos de recombinação. Dentre as regiões, *pol* foi a região com o maior número de sequência recombinantes enquanto em o menor número de sequências recombinantes apresentou recombinação na região *tat*. Além disso, foram identificados dois *hotspots* entre as sequências analisadas: um encontrado na região entre as posições 5360 – 5390 do gene acessório *vif* e outro em torno da posição 9356 da região *nef*.

Foram encontradas assinaturas nucleotídicas conservadas exclusivas de cada CRFs\_BF que podem ser consideradas como um padrão característico de cada CRF permitindo a comparação de sequências, principalmente para auxiliar na determinação de pontos de recombinação em análises futuras.

As mutações mais frequentes foram encontradas na região da transcriptase reversa: M41L (NRTIs) e K100N (NNRTIs). Os recombinantes BF apresentam altas taxas de susceptibilidade aos ARVs DRV, TDF, ETR, RPV e DTG. Já as maiores taxas de resistência foram associadas aos ARVs ATV, SQV, NVP, AZT, D4T, DDI e FTC. No entanto, extrapolações a partir dos resultados observados ou até mesmo correlações destes resultados não foram feitas em função da ausência de dados clínico epidemiológicos suficientes.

De modo geral há uma preferência de uso do coreceptor CXCR5 pelos recombinantes BF. Entretanto, todas as sequências da CRF39 foram identificadas como R5X4/X4, enquanto todas sequências da CRF42 foram classificadas com R5. Foram identificados 23 diferentes

motivos estruturais na laça V3 da proteína gp120, sendo os tetrapeptídeos GPGR (49,4%), GPGQ (20%), APGR (6,7%), GPGK (5,1%), GWGR (4,3%) e GPGG (2%) mais frequentes. Destes, os motivos GPGQ e APGR são preferencialmente R5 e contem o subtipo F na região da alça V3.

## REFERÊNCIAS

- ARTS, E. J.; HAZUDA, D. J. HIV-1 antiretroviral drug therapy. **Cold Spring Harbor perspectives in medicine**, v. 2, n. 4, p. a007161, 2012.
- AULICINO, P. C. et al. Extremely rapid spread of human immunodeficiency virus type 1 BF recombinants in Argentina. **Journal of virology**, v. 81, n. 1, p. 427–429, 2007.
- BELLO, G. et al. Phylodynamics of HIV-1 circulating recombinant forms 12\_BF and 38\_BF in Argentina and Uruguay. **Retrovirology**, v. 7, p. 22, 2010.
- BINKA, M. et al. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. **Journal of virology**, v. 86, n. 1, p. 49–59, 2012.
- BRUSELLES, A. et al. Use of Massive Parallel Pyrosequencing for Near Full-Length Characterization of a Unique HIV Type 1 BF Recombinant Associated with a Fatal Primary Infection. **AIDS Research and Human Retroviruses**, v. 25, n. 9, p. 937–942, 2009.
- CARMONA, R. et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. **Journal of General Virology**, v. 83, n. 1, p. 107–119, 2002.
- CARR, J. K. et al. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. **AIDS (London, England)**, v. 15, n. 15, p. F41-7, 2001.
- CUEVAS, M. T. et al. Short Communication: Biological and Genetic Characterization of HIV Type 1 Subtype B and Nonsubtype B Transmitted Viruses: Usefulness for Vaccine Candidate Assessment. **AIDS Research and Human Retroviruses**, v. 26, n. 9, p. 1019–1025, 2010.
- DE CANDIA, C. et al. Viral replication is enhanced by an HIV-1 intersubtype recombination-derived Vpu protein. **Virology Journal**, v. 7, n. 1, p. 259, 2010.
- DE SÁ FILHO, D. et al. Characterization of the Full-Length Human Immunodeficiency Virus-1 Genome from Recently Infected Subjects in Brazil. **AIDS Research and Human Retroviruses**, v. 23, n. 9, p. 1087–1094, 2007.
- DE SA-FILHO, D. J. et al. HIV type 1 pol gene diversity and antiretroviral drug resistance mutations in Santos, Brazil. **AIDS research and human retroviruses**, v. 24, n. 3, p. 347–353, 2008.
- DELICADO, T. L.; TORRECILLA, E.; HOLGUÍN, ÁFRICA. Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naive individuals. **Journal of Antimicrobial Chemotherapy**, v. 71, n. 2, p. 362–366, 2016.
- FERNÁNDEZ-GARCÍA, A. et al. Identification of a New HIV Type 1 Circulating BF Intersubtype Recombinant Form (CRF47\_BF) in Spain. **AIDS Research and Human Retroviruses**, v. 26, n. 7, p. 827–832, 2010.

FERRARO, G. A. et al. Biological Characterization and Chemokine Receptor Usage of HIV Type 1 Isolates Prevalent in Brazil. **AIDS Research and Human Retroviruses**, v. 17, n. 13, p. 1241–1247, 2001.

GLAZIOU, P. et al. Global Epidemiology of Tuberculosis. **Cold Spring Harbor Perspectives in Medicine**, v. 5, n. 2, p. a017798–a017798, 2015.

GOMEZ CARRILLO, M. et al. Mother-to-child HIV type 1 transmission in Argentina: BF recombinants have predominated in infected children since the mid-1980s. **AIDS Research and Human Retroviruses**, v. 18, n. 7, p. 477–483, 2002.

GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular biology and evolution**, v. 27, n. 2, p. 221–224, 2010.

HALL, T. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT** Nucleic Acids Symposium Series, 1999.

HU, W.; TEMIN, H. Retroviral recombination and reverse transcription. **Science**, v. 250, n. 4985, p. 1227–1233, 1990.

JETZT, A. E. et al. High rate of recombination throughout the human immunodeficiency virus type 1 genome. **Journal of virology**, v. 74, n. 3, p. 1234–40, 2000.

LEAL, É. et al. Evolutionary Dynamics of HIV-1 BF and CB Recombinants and Its Parental Counterparts in South America. **AIDS Research and Human Retroviruses**, n. 55 11, p. 52841–52841, 2008a.

LEAL, É. et al. Molecular and structural characterization of HIV-1 subtype B Brazilian isolates with GWGR tetramer at the tip of the V3-loop. **Virology**, v. 381, n. 2, p. 222–229, 2008b.

MELO, Fernando Lucas de. **Caracterização biológica e molecular de recombinantes naturais de HIV-1**. 2011. Tese (Doutorado em Microbiologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2011. doi:10.11606/T.42.2011.tde-20102011-133503.

MELO, F. L.; JAMAL, L. F.; ZANOTTO, P. M. D. A. Characterization of primary isolates of HIV type 1 CRF28\_BF, CRF29\_BF, and unique BF recombinants circulating in São Paulo, Brazil. **AIDS research and human retroviruses**, v. 28, n. 9, p. 1082–1088, 2012.

MICHAEL M. THOMSON et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. **Journal of General Virology**, v. 83, n. 2002, p. 107–119, 2002.

Protocolo clínico e diretrizes terapêuticas para manejo da infecção pelo HIV em adultos (atualizado em 31/07/2015). Ministério da Saúde. Brasília, 2013.

PESSOA, R. et al. Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72\_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil. **Genome Announcements**, v. 2, n. 3, p. e00386-14-e00386-14, 2014a.

PESSÔA, R. et al. Deep Sequencing of HIV-1 near Full-Length Proviral Genomes Identifies High Rates of BF1 Recombinants Including Two Novel Circulating Recombinant Forms (CRF) 70\_BF1 and a Disseminating 71\_BF1 among Blood Donors in Pernambuco, Brazil. **PLoS ONE**, v. 9, n. 11, p. e112674, 2014b.

PESSÔA, R. et al. Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus Type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. **Transfusion**, v. 55, n. 5, p. 980–990, 2015.

PESSÔA, R. et al. Ultra-Deep Sequencing of HIV-1 near Full-Length and Partial Proviral Genomes Reveals High Genetic Diversity among Brazilian Blood Donors. **PLOS ONE**, v. 11, n. 3, p. e0152499, 2016.

POSADA, D.; CRANDALL, K. A. Bioinformatics Applications Note Modeltest : testing the model of DNA substitution. **Bioinformatics applications note**, v. 14, n. 9, p. 817–818, 1998.

ROBERTSON, D. L. HIV-1 Nomenclature Proposal. **Science**, v. 288, n. 5463, p. 55, 2000.

SABINO, E. C. et al. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. **Journal of virology**, v. 68, n. 10, p. 6340–6346, 1994.

SALMINEN, M. O. et al. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. **AIDS Research and Human Retroviruses**, v. 11, n. 11, p. 1423–1425, 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 São Paulo, Brazil. **Infection, Genetics and Evolution**, v. 6, n. 5, p. 368–377, 2006.

SANABANI, S. et al. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. **Virology Journal**, v. 7, n. 1, p. 74, 2010.

SANABANI, S. S. et al. Characterization of Partial and Near Full-Length Genomes of HIV-1 Strains Sampled from Recently Infected Individuals in São Paulo, Brazil. **PLoS ONE**, v. 6, n. 10, p. e25869, 2011.

SANABANI, S. S. et al. Variability of HIV-1 Genomes among Children and Adolescents from São Paulo, Brazil. **PLoS ONE**, v. 8, n. 5, p. e62552, 2013.

SANTORO, M. M.; PERNO, C. F. HIV-1 Genetic Variability and Clinical Implications. **ISRN microbiology**, v. 2013, n. August, p. 1-20, 2013.

SIERRA, M. et al. The analysis of near full-length genome sequences of human

immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12\_BF. **Infection, Genetics and Evolution**, v. 5, n. 3, p. 209–217, 2005a.

SIERRA, M. et al. The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12\_BF. **Infection, Genetics and Evolution**, v. 5, n. 3, p. 209–217, 2005b.

SIMONETTI, F. R. et al. Identification of a new HIV-1 BC circulating recombinant form (CRF60\_BC) in Italian young men having sex with men. **Infection, Genetics and Evolution**, v. 23, p. 176–181, 2014.

SWOFFORD, D. L. Phylogenetic Analysis Using Parsimony. **Options**, v. 42, n. 2, p. 294–307, 2002.

THOMSON, M. M. et al. Analysis of near full-length genome sequences of HIV type 1 BF intersubtype recombinant viruses from Brazil reveals their independent origins and their lack of relationship to CRF12\_BF. **AIDS research and human retroviruses**, v. 20, n. 10, p. 1126–33, 2004.

UNAIDS. **Fact Sheet 2016 Global Statistics** — 2015. p. 1–8, 2016.

VALLARI, A. et al. Confirmation of Putative HIV-1 Group P in Cameroon. **Journal of Virology**, v. 85, n. 3, p. 1403–1407, 2011.

VILLANOVA, F. E. Diversity of HIV-1 Subtype B : Implications to the Origin of BF Recombinants. v. 5, n. 7, p. 1–9, 2010.

WHO. **Global update on the health sector response to HIV**, 2014. Disponível em: <<http://www.who.int/hiv/pub/progressreports/update2014/en/>>.

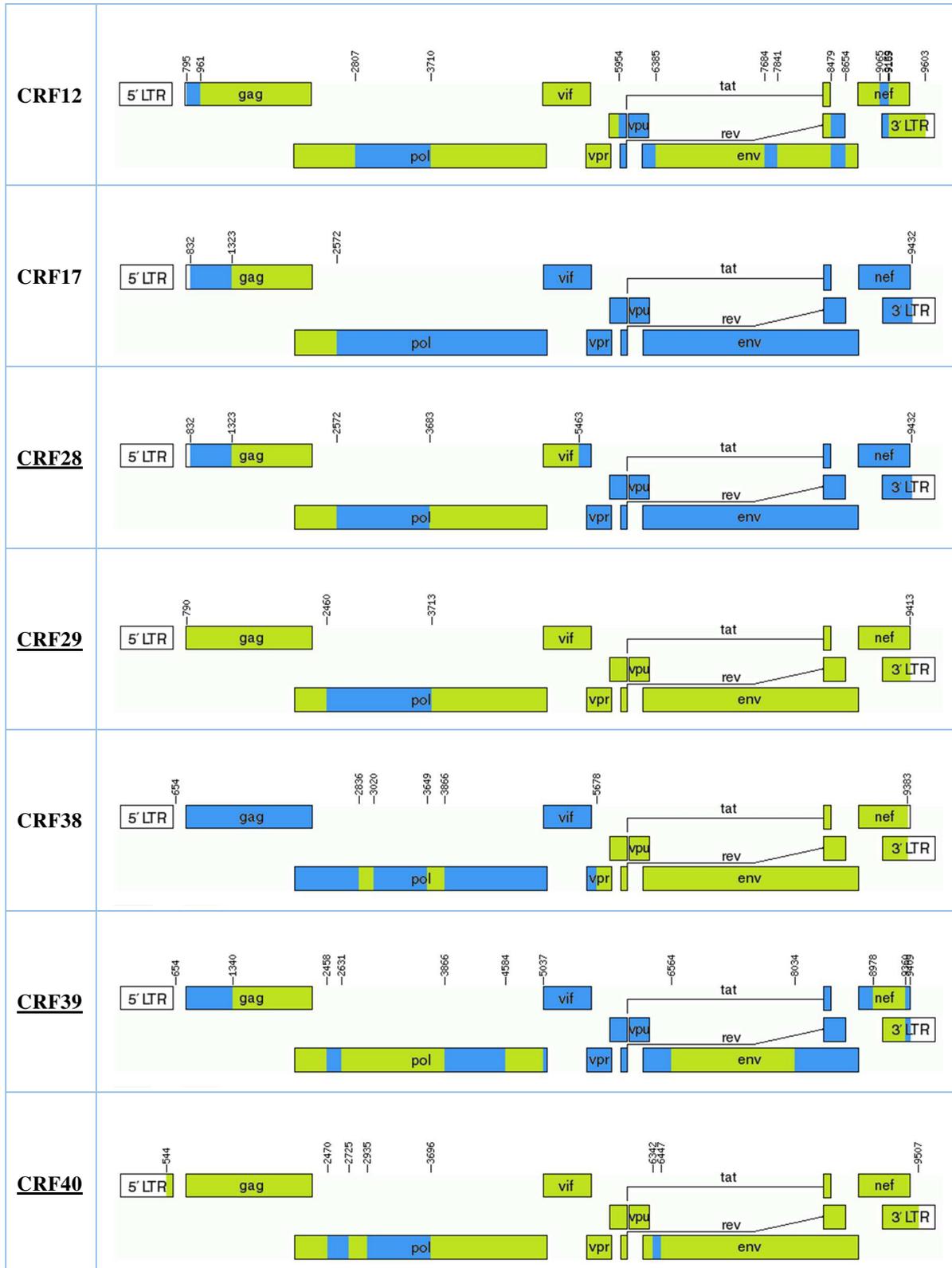
ZHANG, M. TRACKING HIV-1 GENETIC VARIATION : 2007.

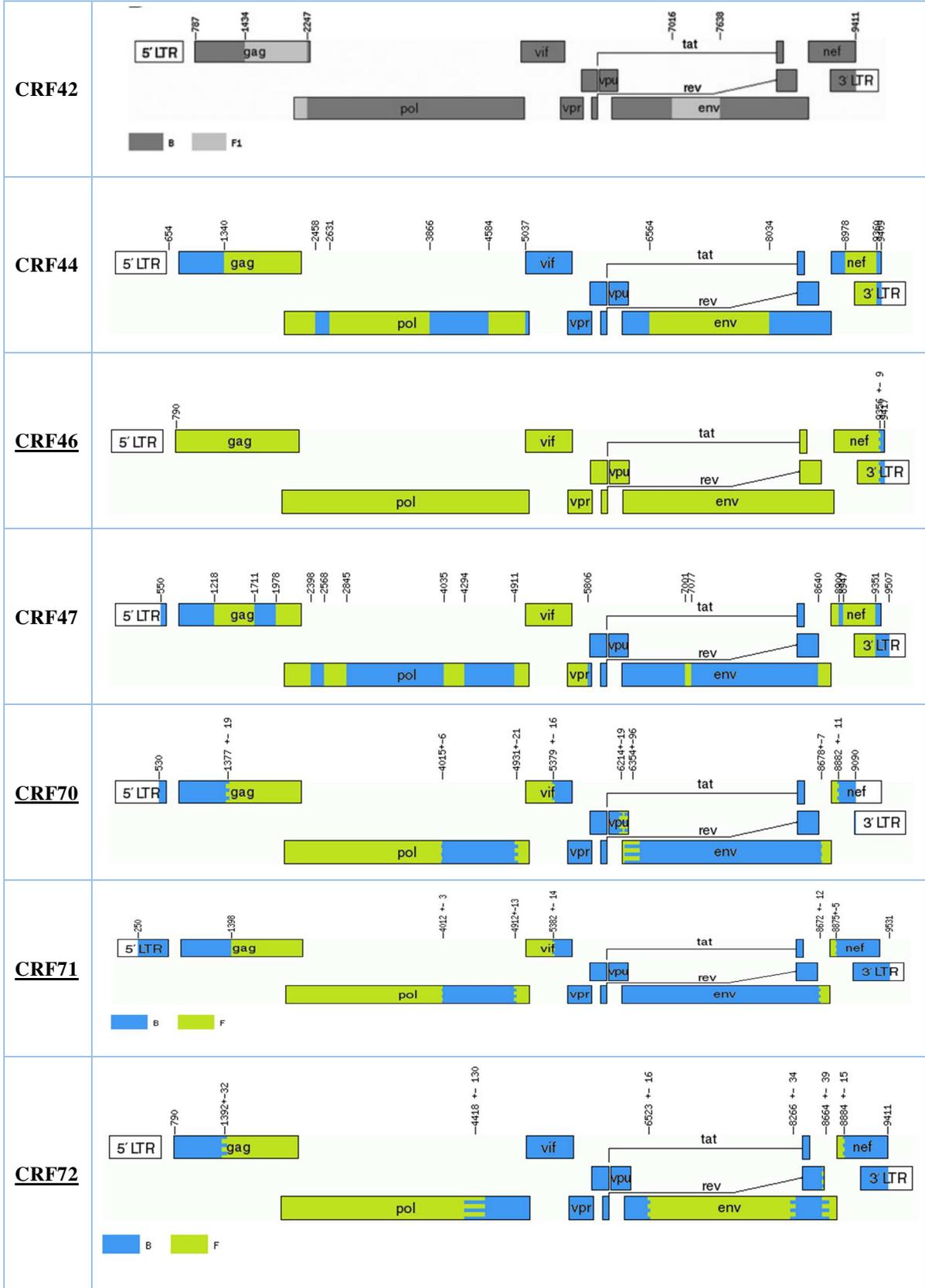
ZHANG, Z. et al. Vif Proteins from Diverse Human Immunodeficiency Virus / Simian Immunodeficiency Virus Lineages Have Distinct Binding Sites in A3C. v. 90, n. 22, p. 10193–10208, 2016.

**APÊNDICE A** - Tabela com números de acesso no GeneBank das sequências utilizadas neste estudo.

Número de acesso do GeneBank									
AB253430	AY037277	AY771593	DQ383755	EU581824	GU595150	JF804812	JX140661	KJ849777	KT427690
AB253431	AY037278	AY781128	EF637052	EU581825	GU595151	JF804813	KC113006	KJ849778	KT427698
AB480298	AY037279.2	DQ085867	EF637055	EU581826	GU595152	JF804814	KC113008	KJ849779	KT427705
AB480299	AY037280	DQ085868	EU170135	EU581827	GU595153	JN235952	KC473839	KJ849783	KT427710
AB480300	AY037281	DQ085869	EU170136	EU581828	GU595154	JN235953	KC473840	KJ849787	KT427722
AB480301	AY037283	DQ085870	EU170137	EU735534	GU595155	JN235954	KJ671533	KJ849791	KT427746
AF005495	AY455778	DQ085871	EU170138	EU735535	GU595156	JN235955	KJ671534	KJ849794	KT427758
AF332867	AY455779	DQ085872	EU170139	EU735536	GU595158	JN235956	KJ671535	KJ849797	KT427759
AF385934	AY455780	DQ085873	EU170140	EU735537	GU595159	JN235960	KJ671536	KJ849798	KT427774
AF385935	AY455781	DQ085874	EU170141	EU735538	GU595160	JN235962	KJ671537	KJ849800	KT427776
AF385936	AY455782	DQ085875	EU170142	EU735539	GU595161	JN235963	KJ849757	KJ849802	KT427779
AF408626	AY455783	DQ085876	EU170143	EU735540	GU595162	JN235964	KJ849758	KJ849809	KT427780
AF408627	AY455784	DQ358799	EU170144	FJ213780	HM026455	JN251897	KJ849759	KJ849810	KT427782
AF408628	AY455785	DQ358800	EU170145	FJ213781	HM026456	JN251898	KJ849760	KJ849813	KT427789
AF408629	AY536233	DQ358801	EU170146	FJ213782	HM026457	JN251899	KJ849761	KJ849816	KT427793
AF408630	AY536234	DQ358802	EU170147	FJ213783	HM026458	JN251900	KJ849762	KJ849826	KT427795
AF408631	AY536235	DQ358803	EU170148	FJ358521	HM026459	JN251902	KJ849768	KT276259	KT427798
AF408632	AY536236	DQ358804	EU170149	FJ670529	HM026460	JN251903	KJ849769	KT276269	KT427801
AY037266	AY536237	DQ358806	EU170150	FJ853621	JF804805	JN251904	KJ849770	KT276271	KT427807
AY037267	AY536238	DQ358807	EU170151	FJ904244	JF804806	JN251905	KJ849771	KT427649	KT427809
AY037271	AY771588	DQ358811	EU170152	GQ372987	JF804807	JN692437	KJ849772	KT427650	KT427816
AY037272	AY771589	DQ358812	EU170153	GQ372989	JF804808	JN692448	KJ849773	KT427651	KT427820
AY037273	AY771590	DQ383747	EU170154	GU207884	JF804809	JN692449	KJ849774	KT427652	KT427823
AY037275	AY771591	DQ383753	EU170155	GU595148	JF804810	JN692456	KJ849775	KT427668	KT427825
AY037276	AY771592	DQ383754	EU581823	GU595149	JF804811	JX140660	KJ849776	KT427672	KT427828
								KT427679	KT427871

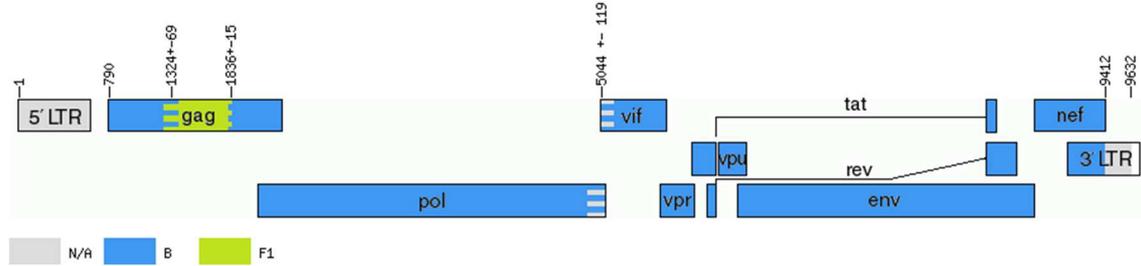
**APÊNDICE B** - Padrões de recombinação das CRFs\_BF descritas até o momento. As CRFs sublinhadas foram descritas no Brasil.



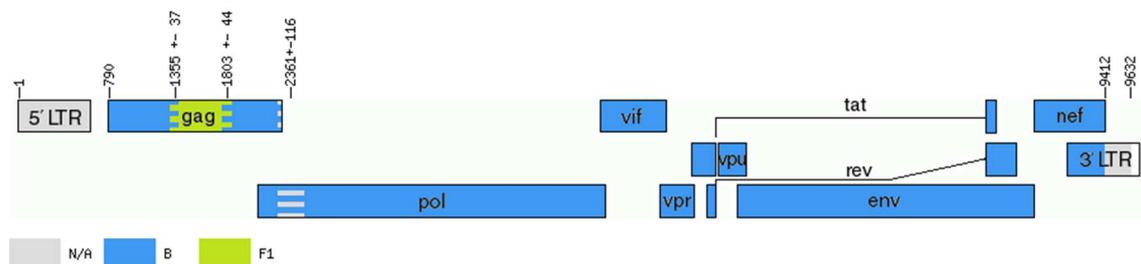


APÊNDICE C – Padrões de recombinação identificados no jpHMM.

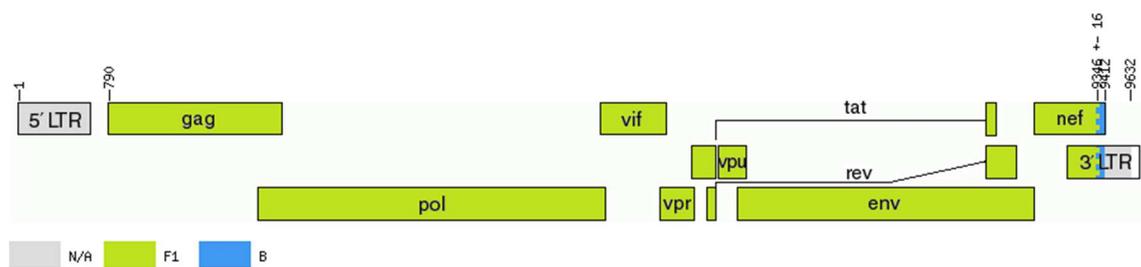
**AB253430.JP.ND.BF1**



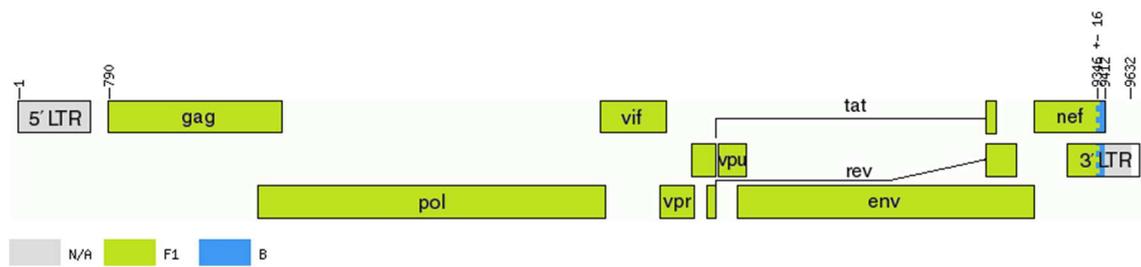
**AB253431.JP.ND.BF1**



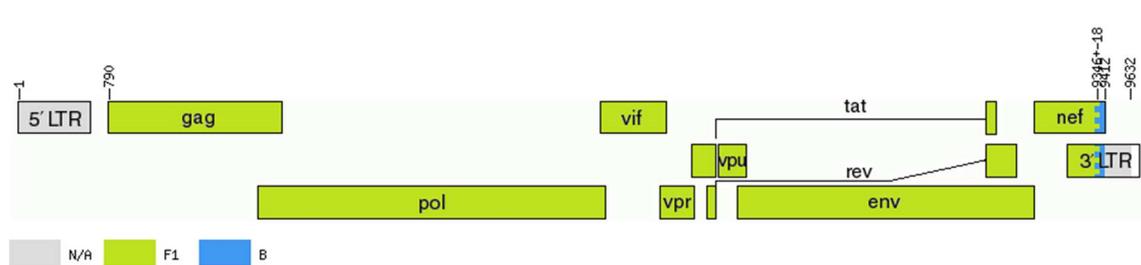
**AB480298.JP.2004.BF1**



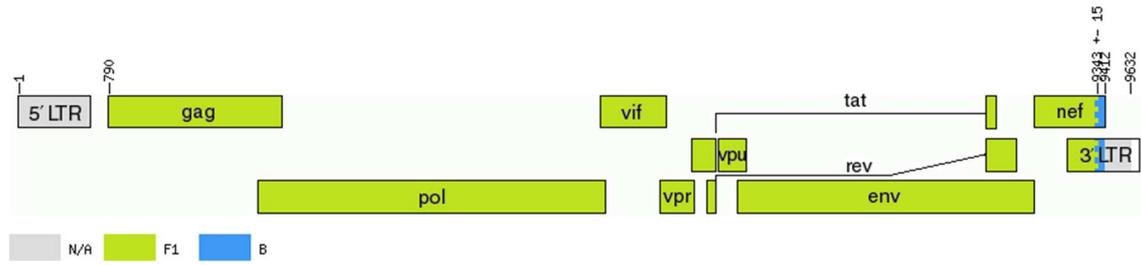
**AB480299.JP.2004.BF1**



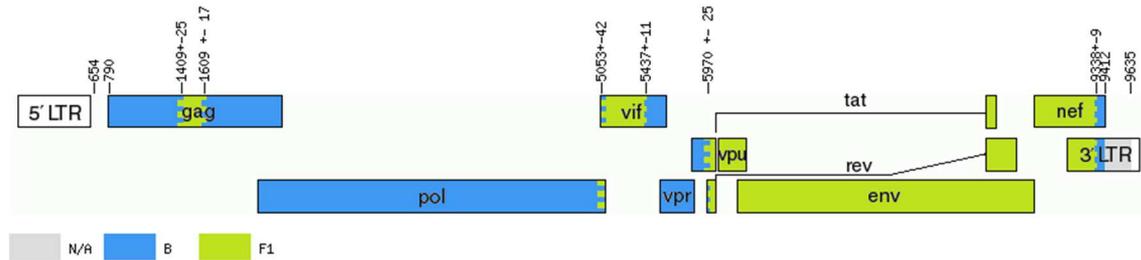
**AB480300.JP.2004.BF1**



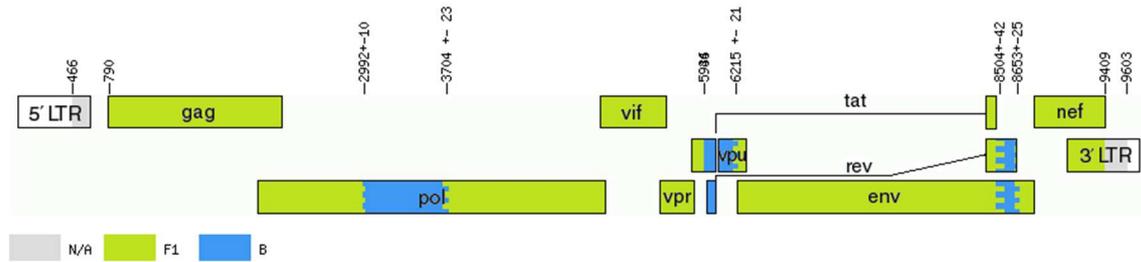
### AB480301.JP.2004.BF1



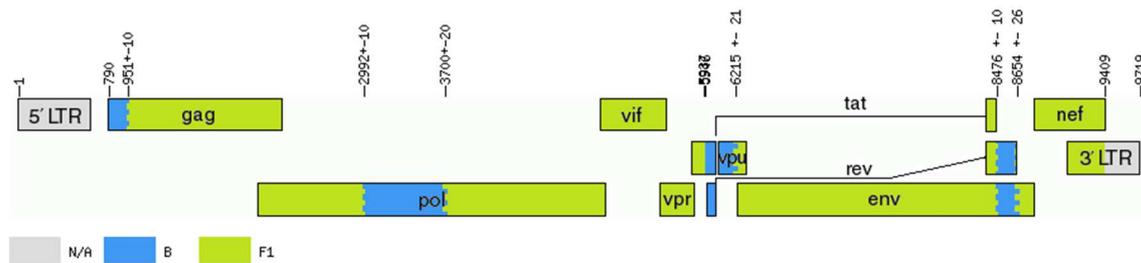
### AF005495.BR.1993.BF1



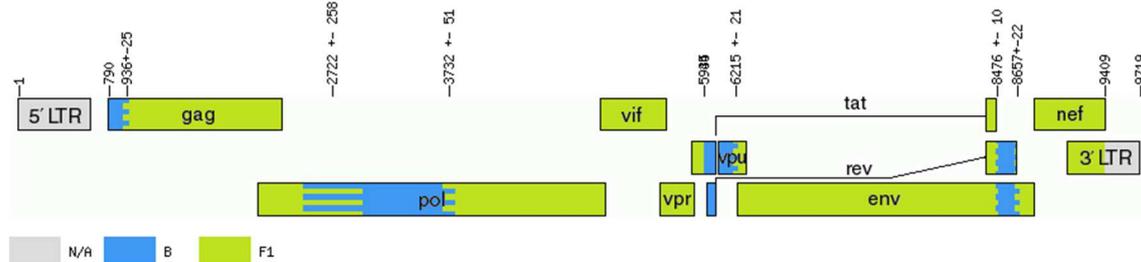
### AF332867.AR.1999.BF



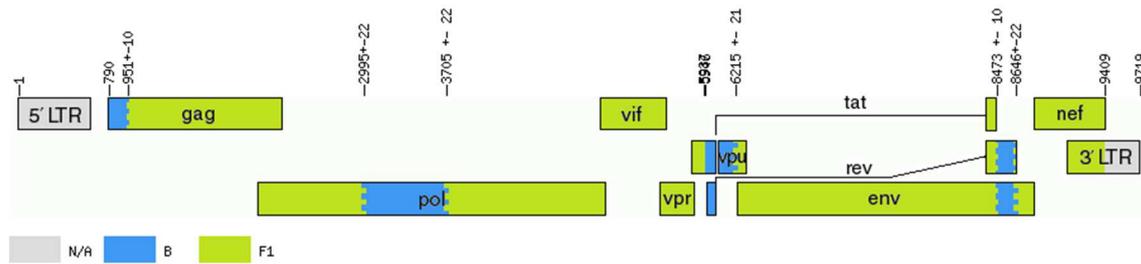
### AF385934.UR.1999.12



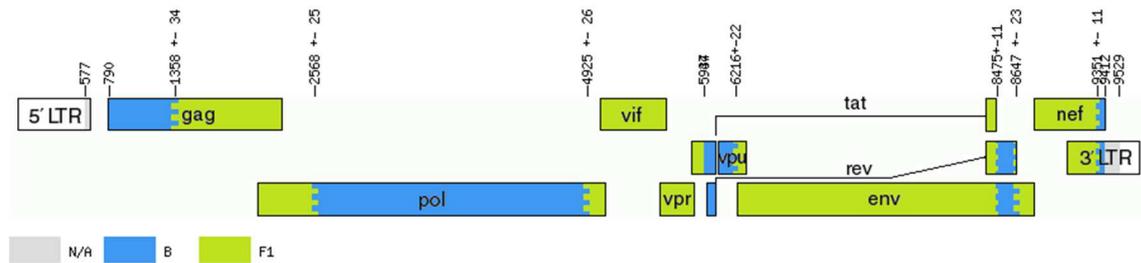
### AF385935.UR.1999.12



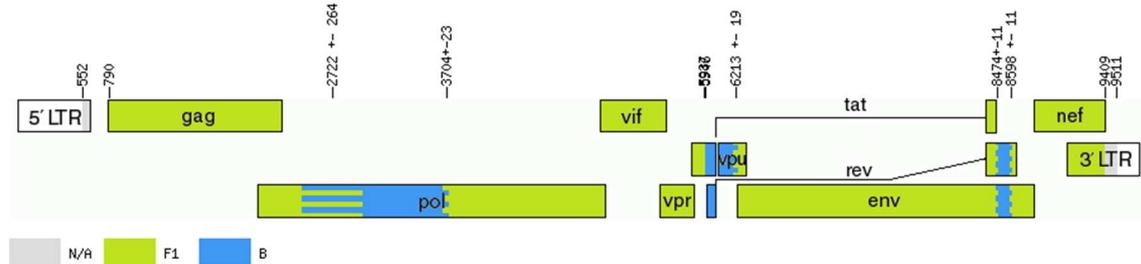
## AF385936.AR.1999.12



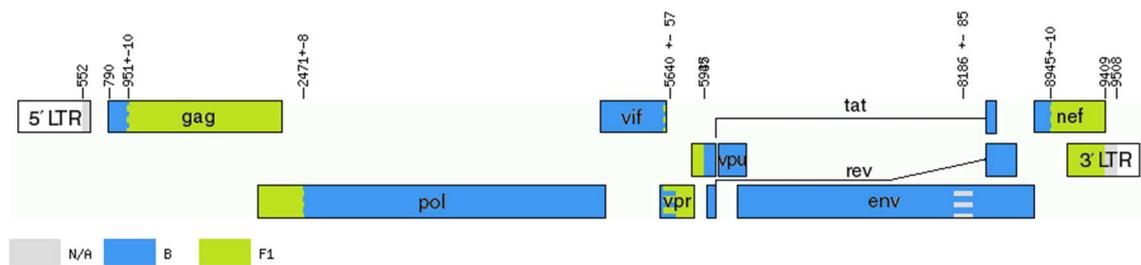
## AF408626.AR.1999.BF



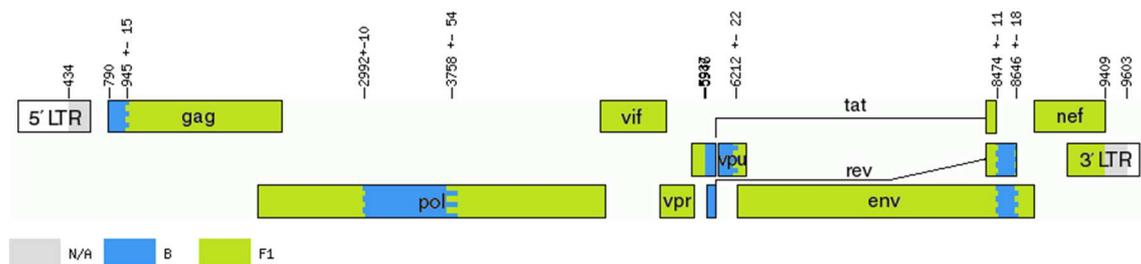
## AF408627.AR.1999.BF



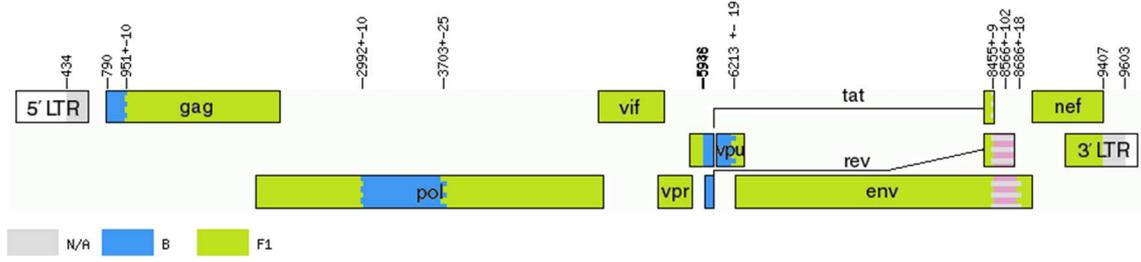
## AF408628.AR.1999.BF



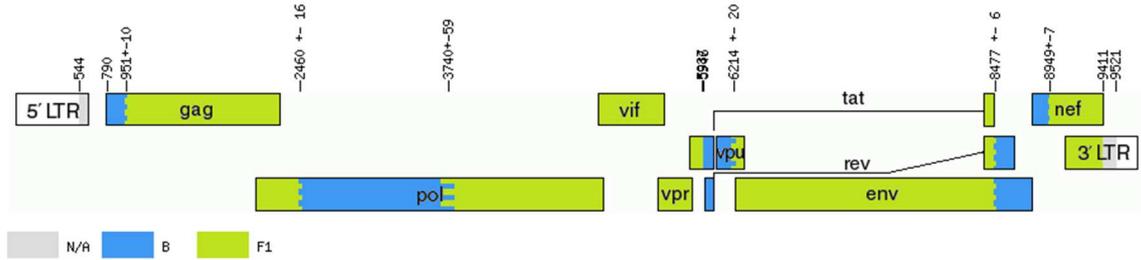
## AF408629.AR.1997.12



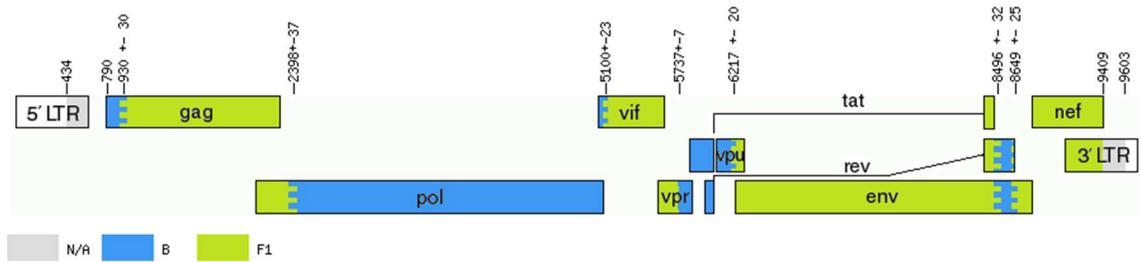
**AF408630.AR.1997.12**



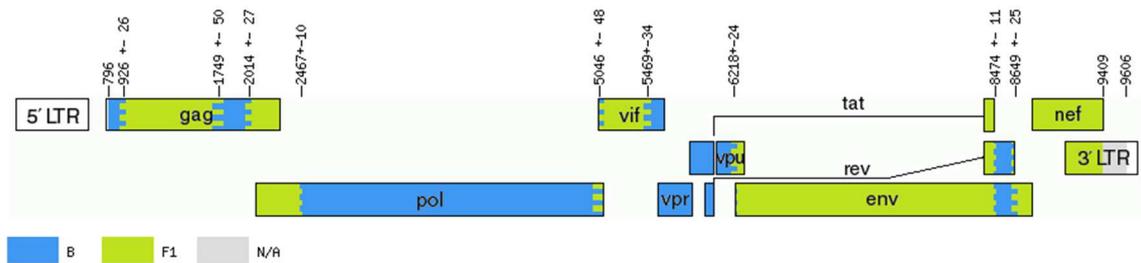
**AF408631.AR.1999.BF**



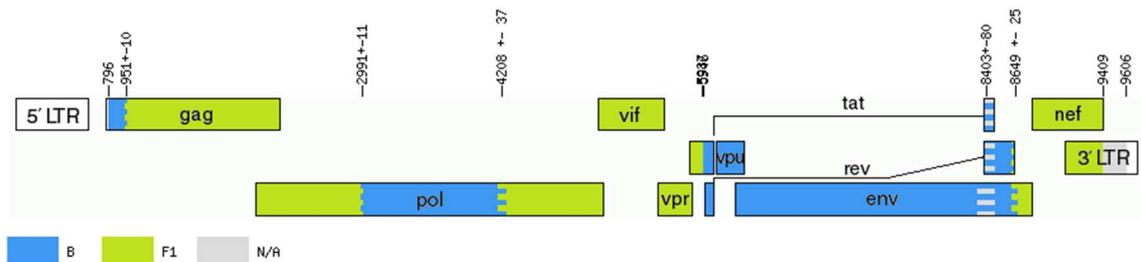
**AF408632.AR.1997.BF**



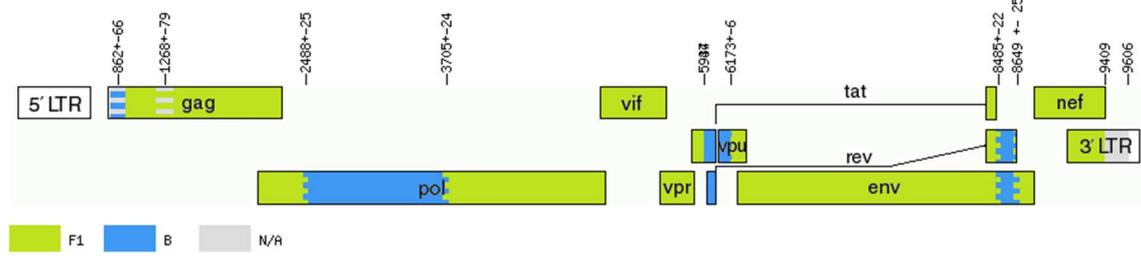
**AY037266.AR.1998.BF**



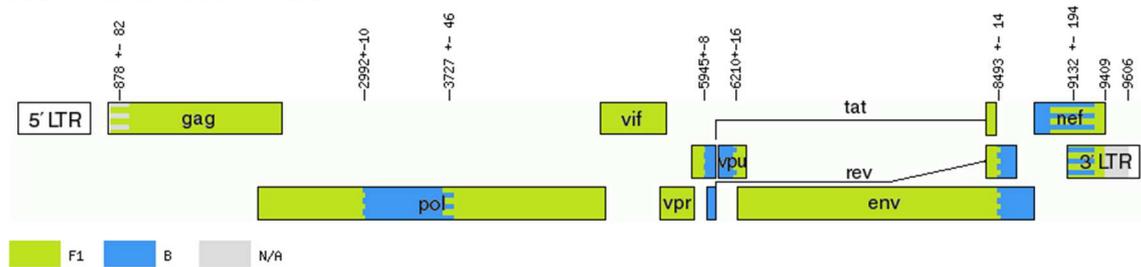
**AY037267.AR.2000.BF**



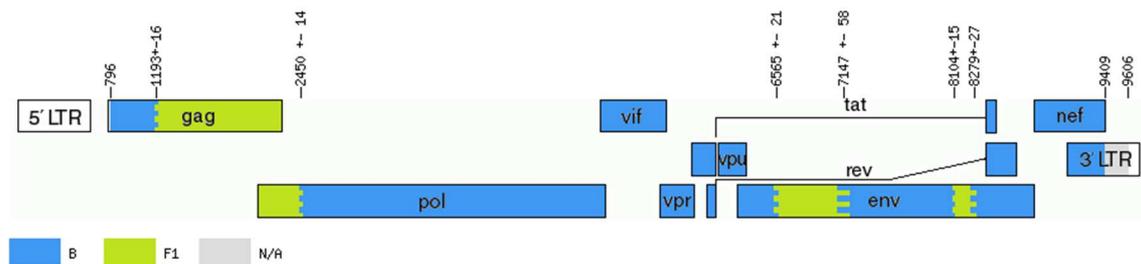
## AY037271.BO.1999.BF



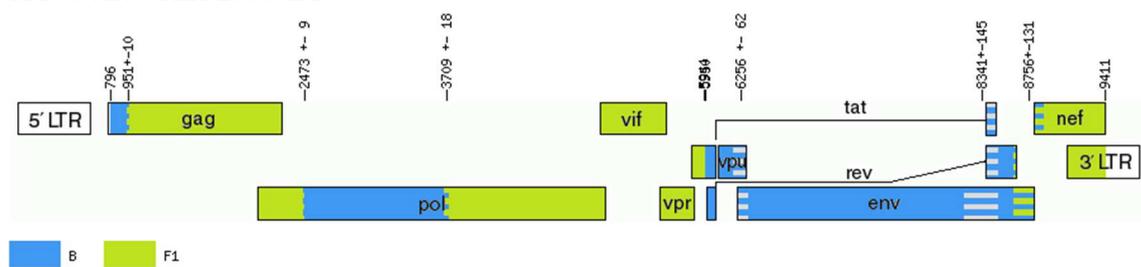
## AY037272.UR.1999.BF



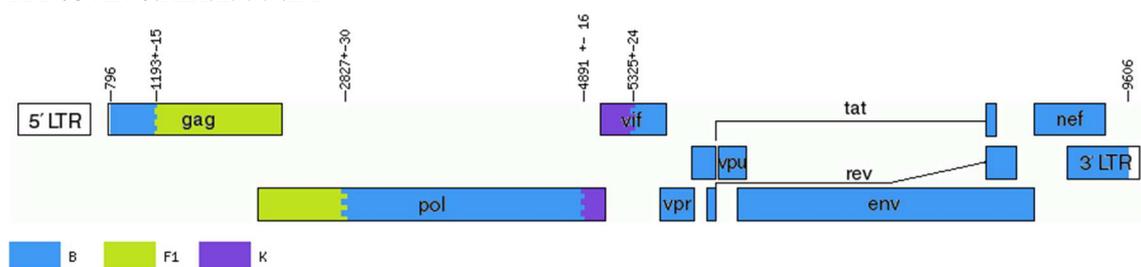
## AY037273.AR.1999.BF



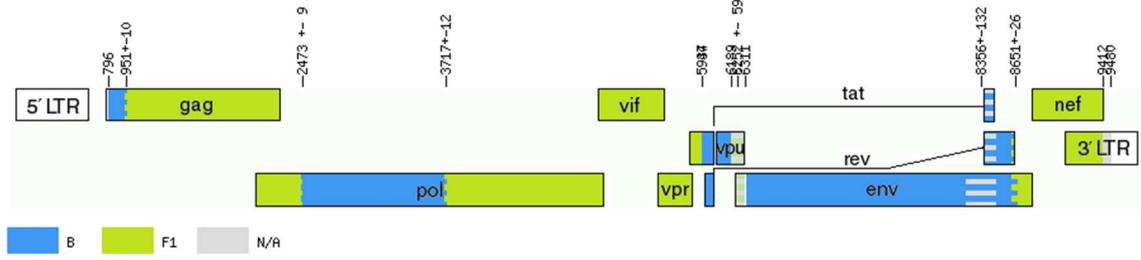
## AY037275.AR.1999.BF



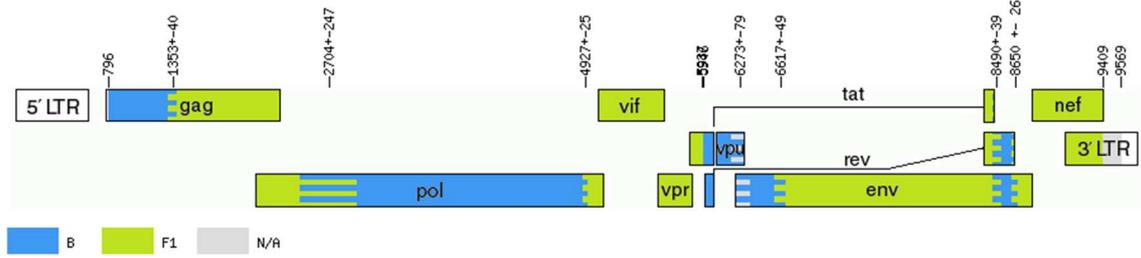
## AY037276.AR.1999.BF



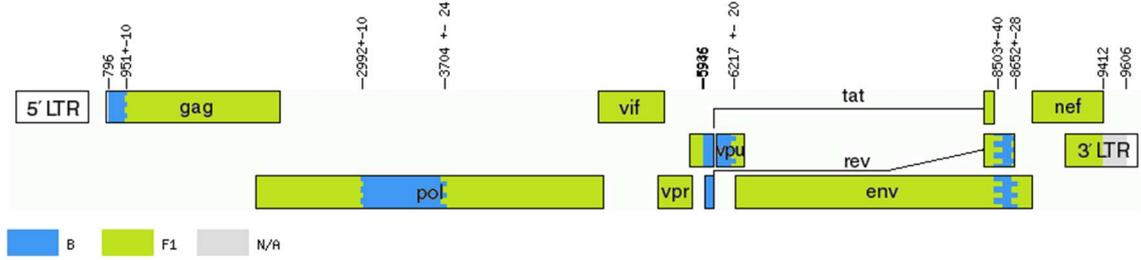
**AY037277.AR.1999.BF**



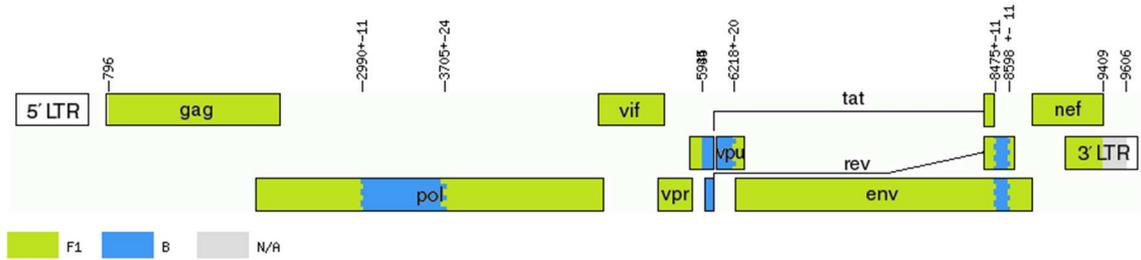
**AY037278.AR.1999.BF**



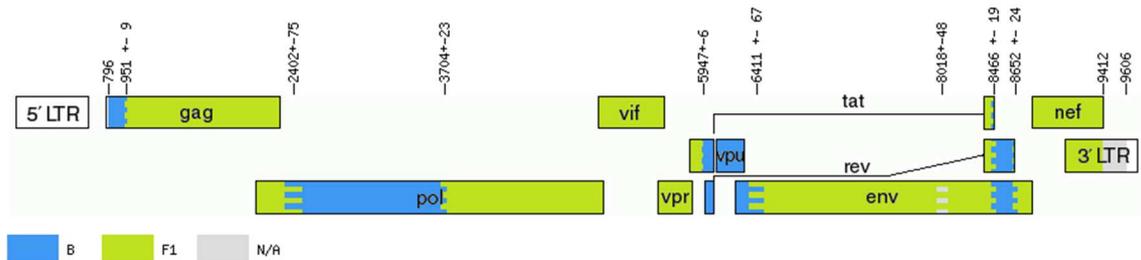
**AY037279.2.AR.1999.12**

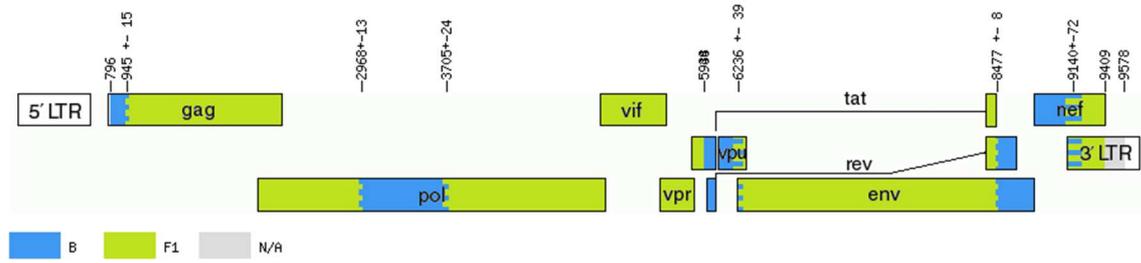
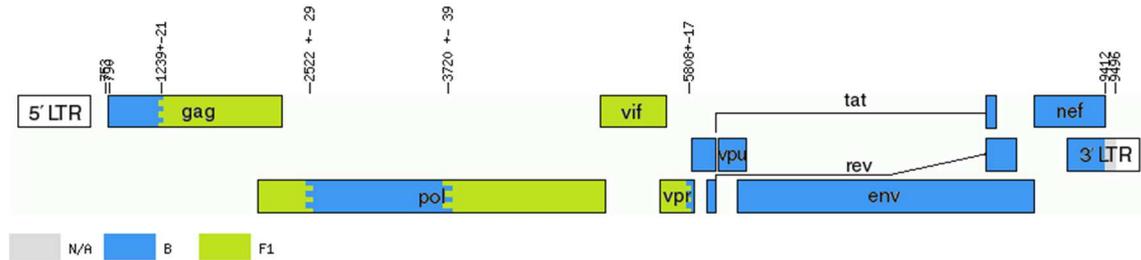
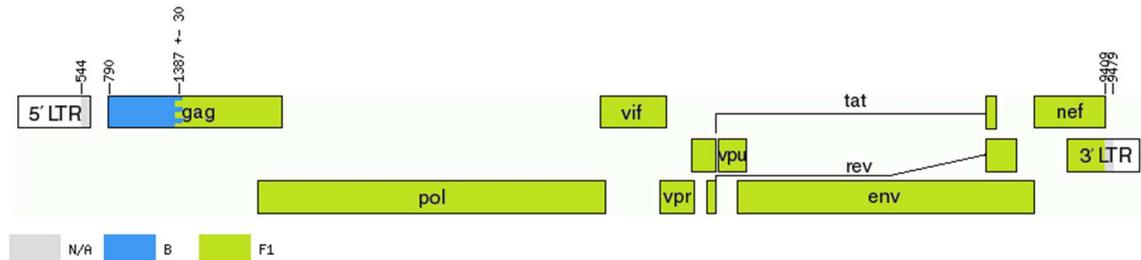
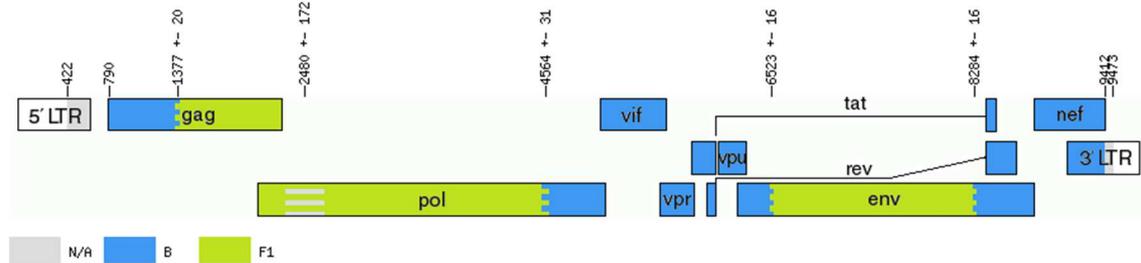
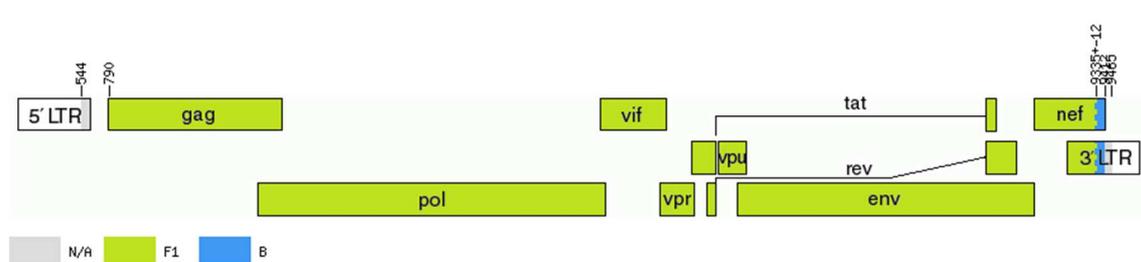


**AY037280.AR.1999.BF**

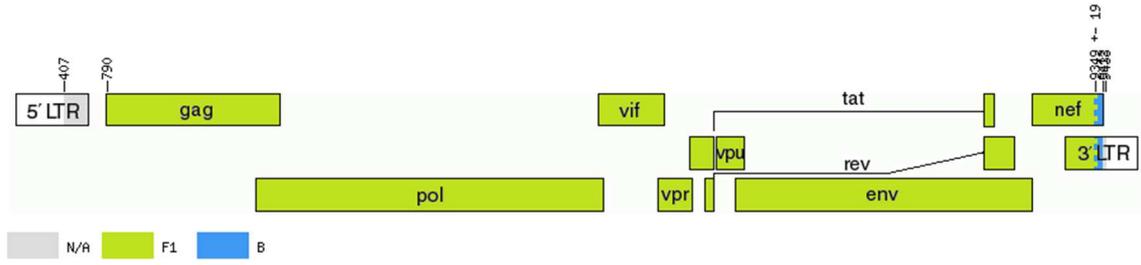


**AY037281.AR.1999.17**

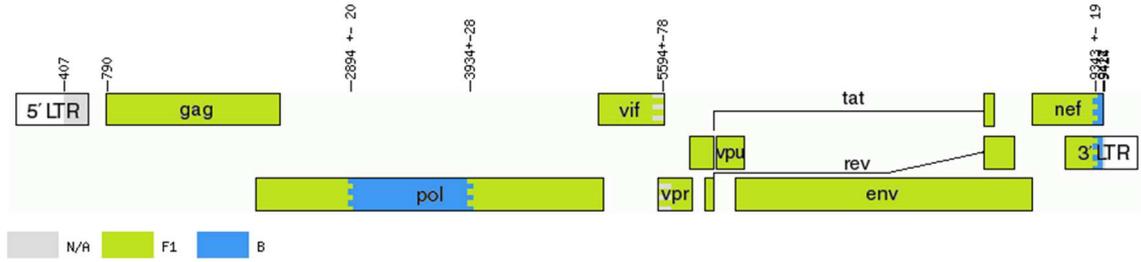


**AY037283.AR.1999.BF****AY455778.BR.1999.29****AY455779.BR.1994.BF****AY455780.BR.1999.BF****AY455781.BR.1994.BF1**

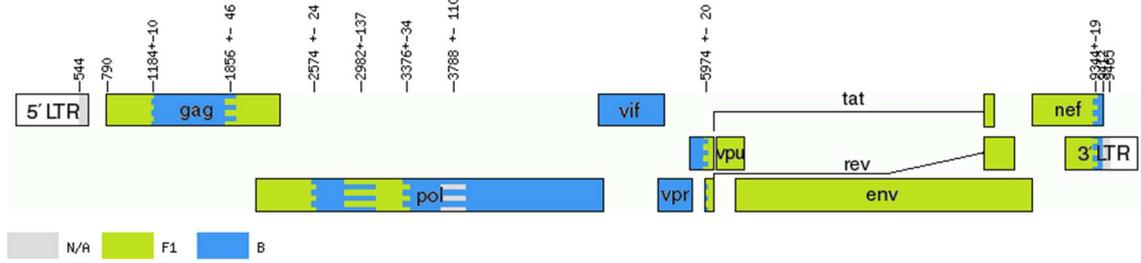
**AY455782.BR.1999.BF1**



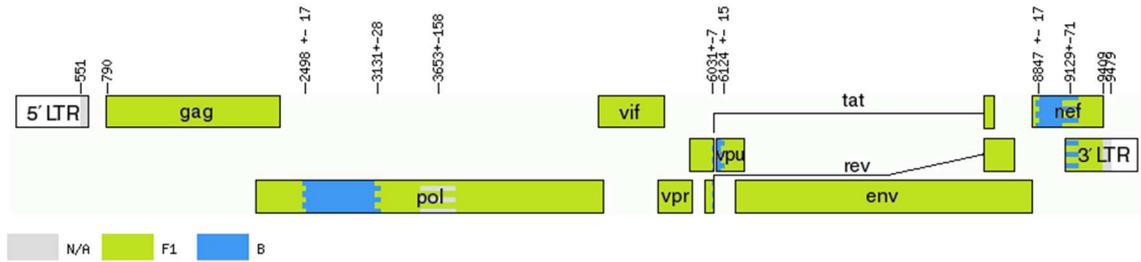
**AY455783.BR.1999.BF**



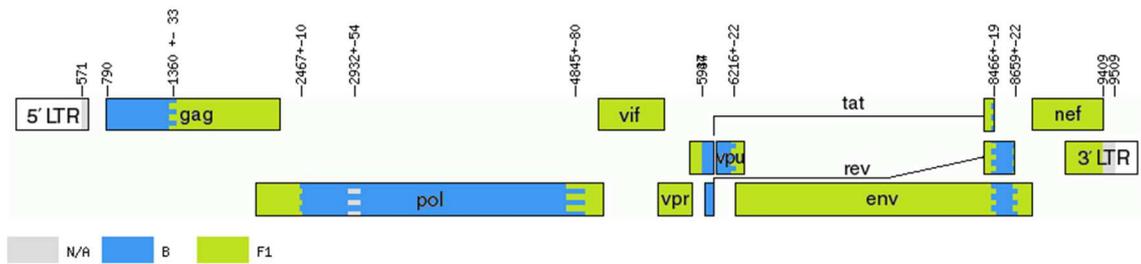
**AY455784.BR.1994.BF**



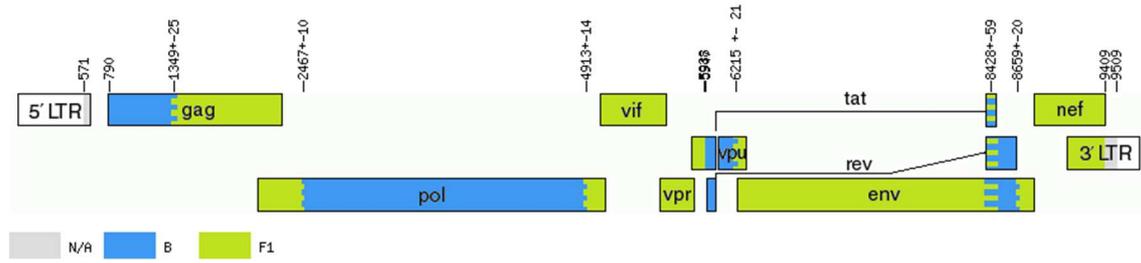
**AY455785.BR.1994.BF**



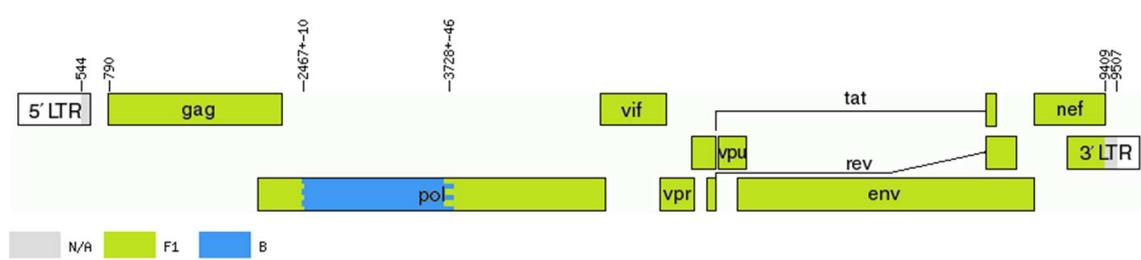
**AY536233.CH.2001.BF**



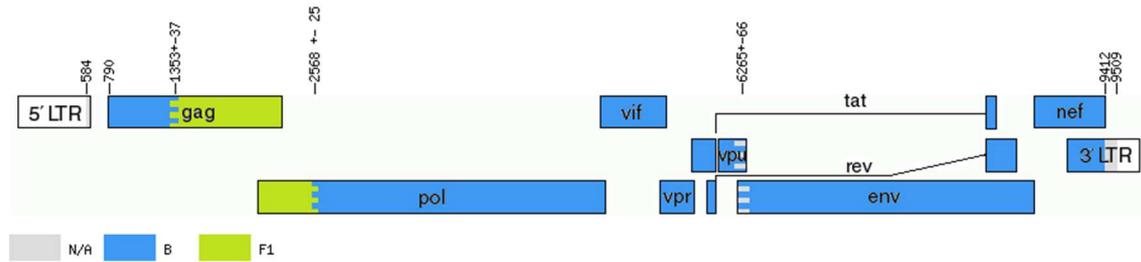
### AY536234.CH.2001.BF



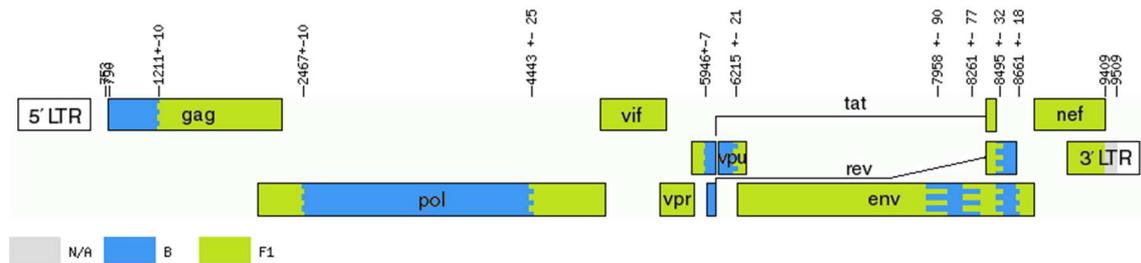
### AY536235.CH.2001.44



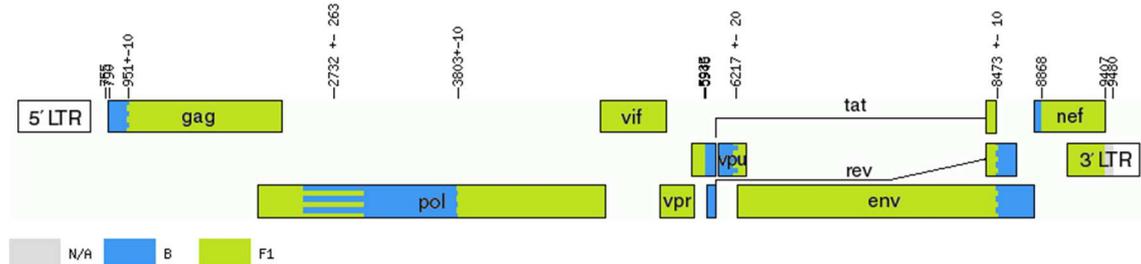
### AY536236.VZ.1999.BF1



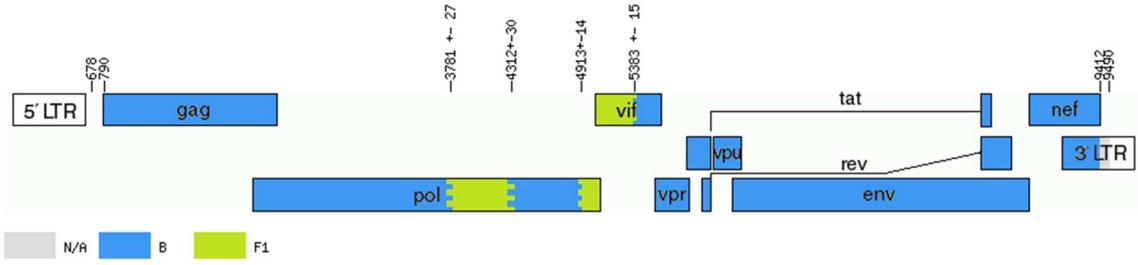
### AY536237.ES.2002.BF



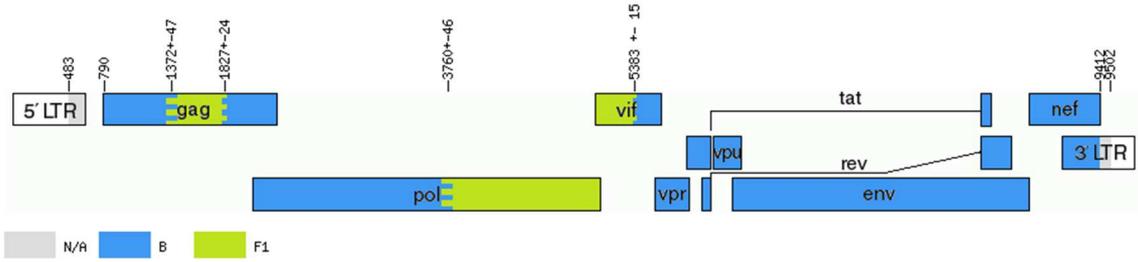
### AY536238.ES.2002.BF1



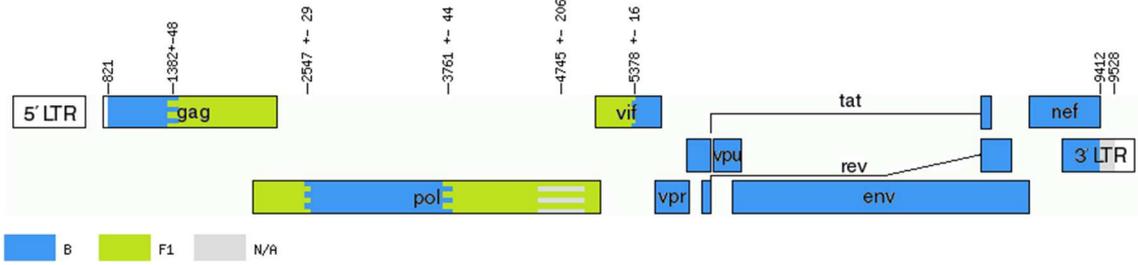
**AY771588.BR.1999.BF1**



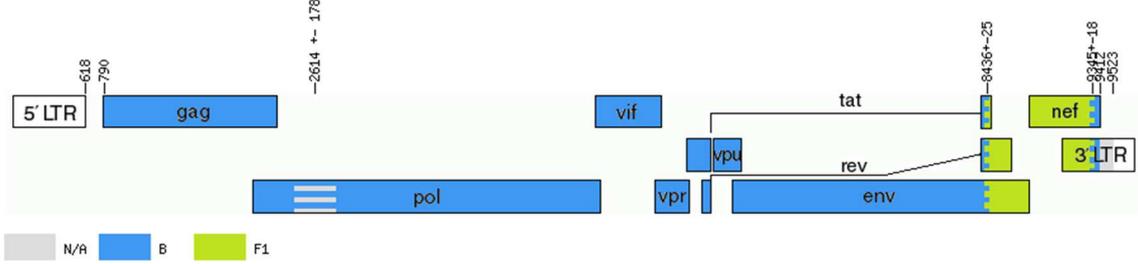
**AY771589.BR.1999.BF1**



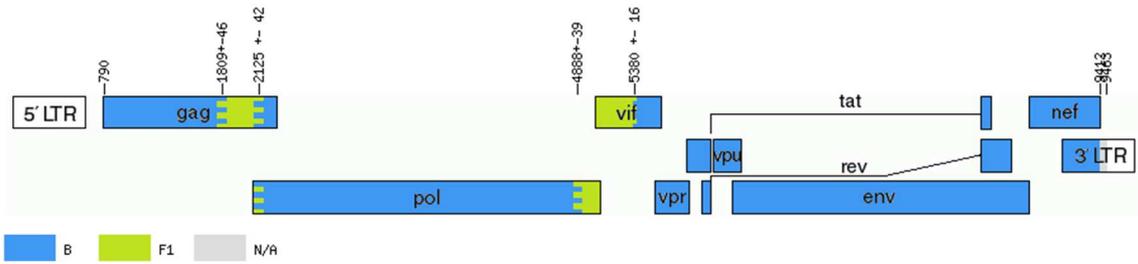
**AY771590.BR.2002.BF**



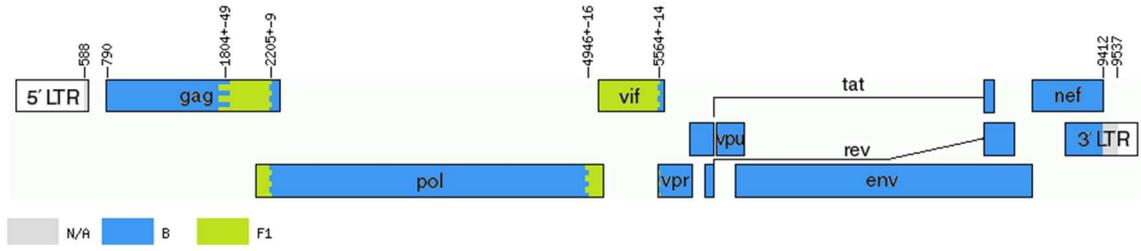
**AY771591.BR.1999.BF**



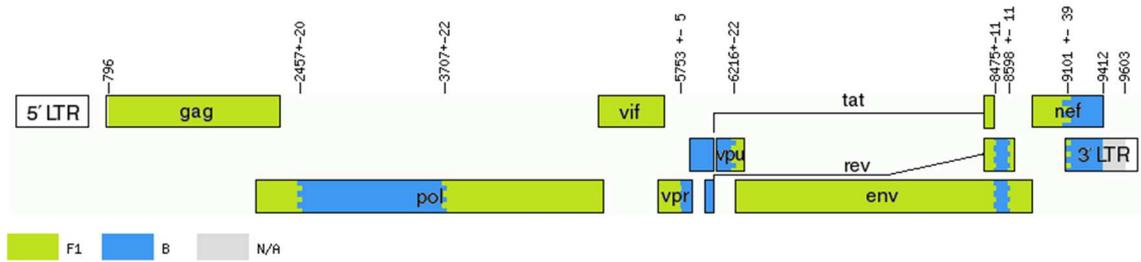
**AY771592.BR.1999.BF**



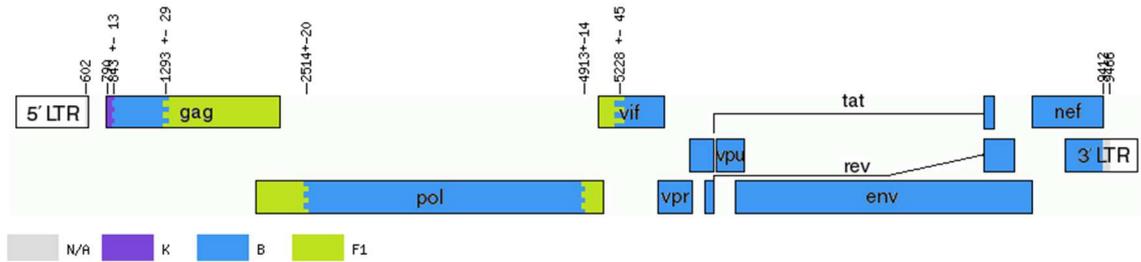
**AY771593.BR.1999.BF**



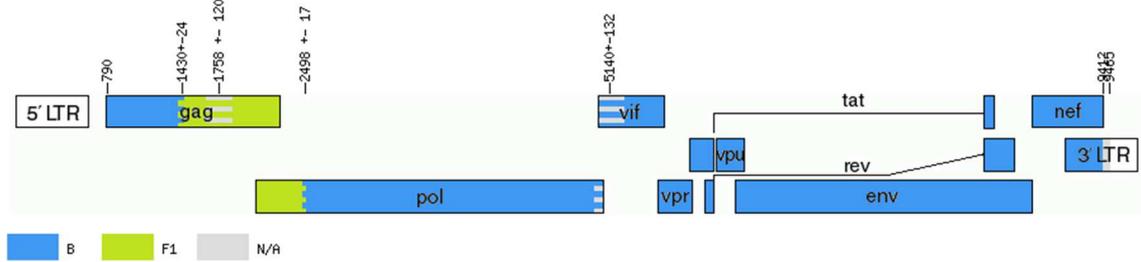
**AY781128.UR.2001.BF**



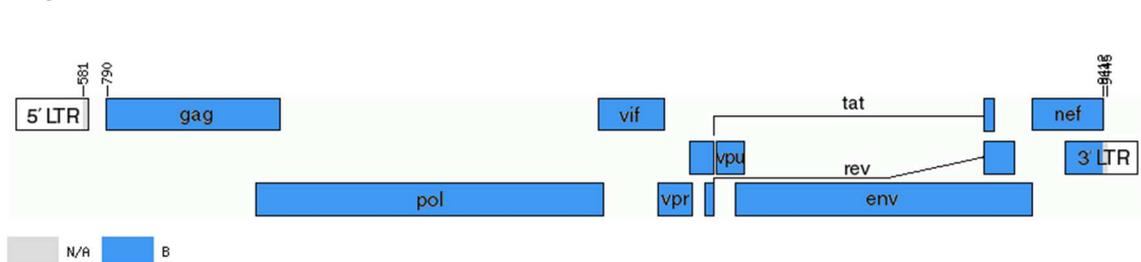
**DQ085867.BR.1999.BF1**



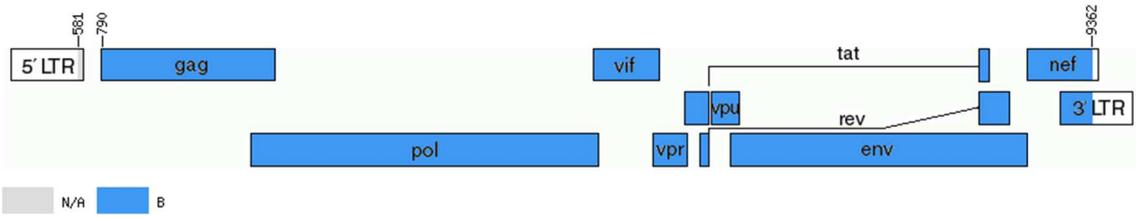
**DQ085868.BR.1999.BF1**



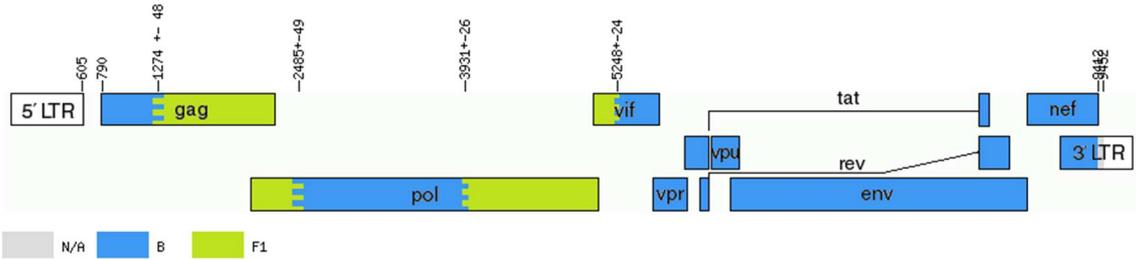
**DQ085869.BR.1999.BF1**



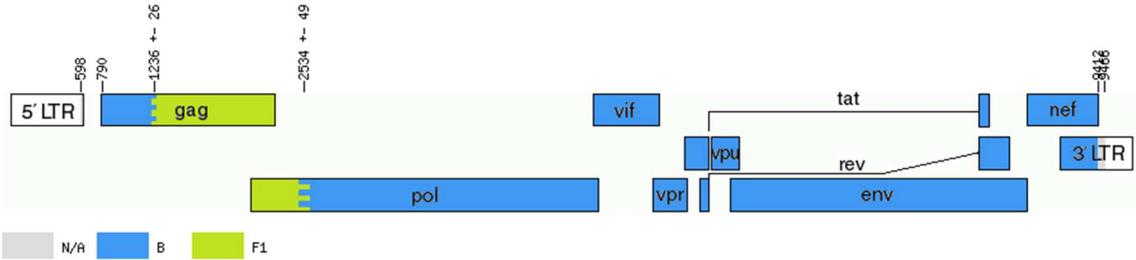
**DQ085870.BR.1999.BF1**



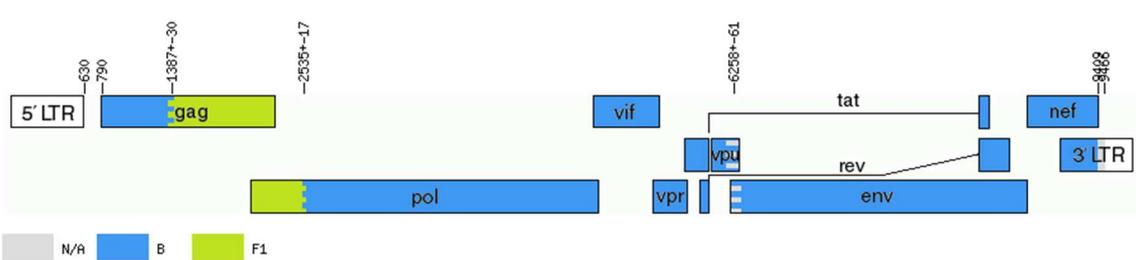
**DQ085871.BR.1999.29**



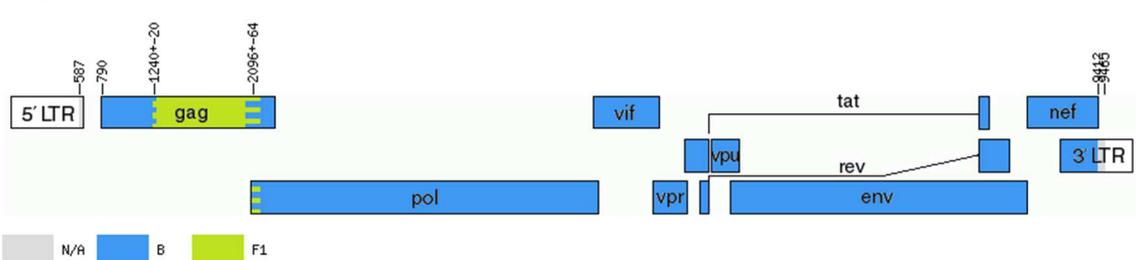
**DQ085872.BR.1999.28**

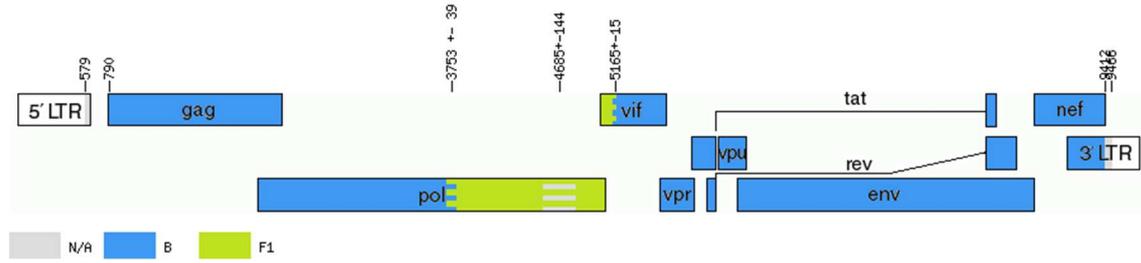
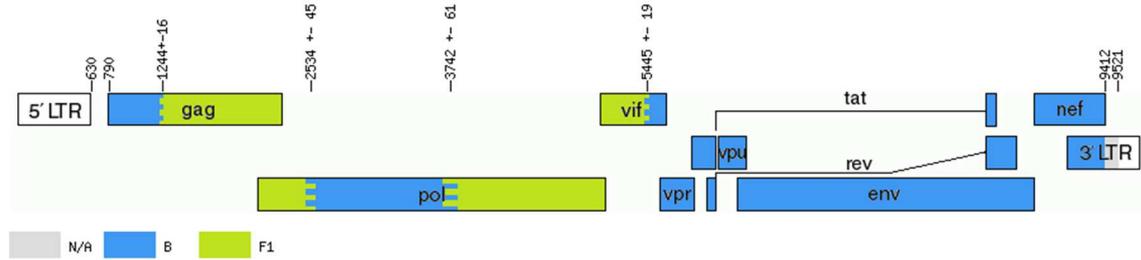
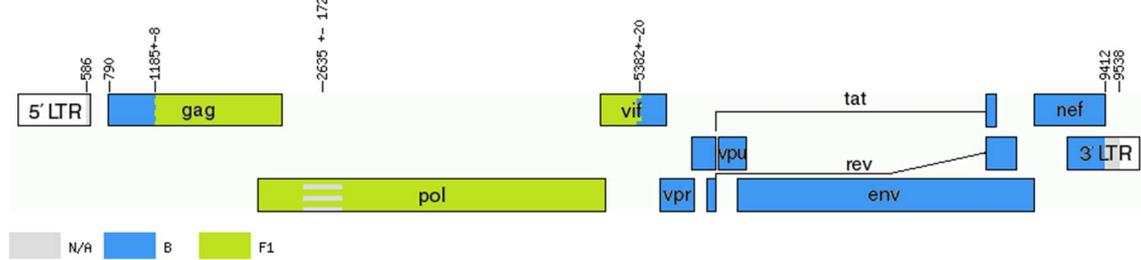
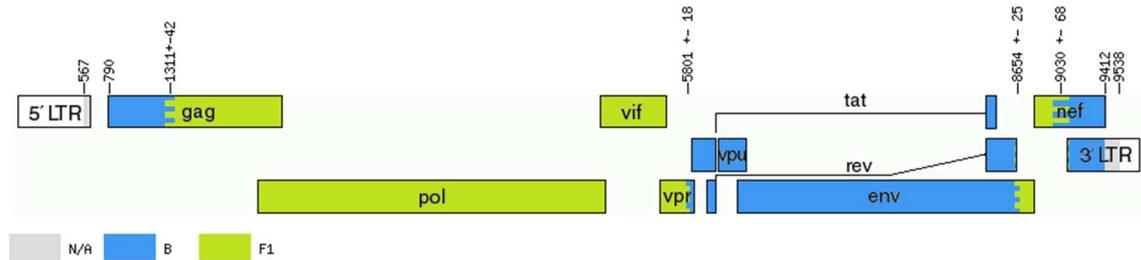
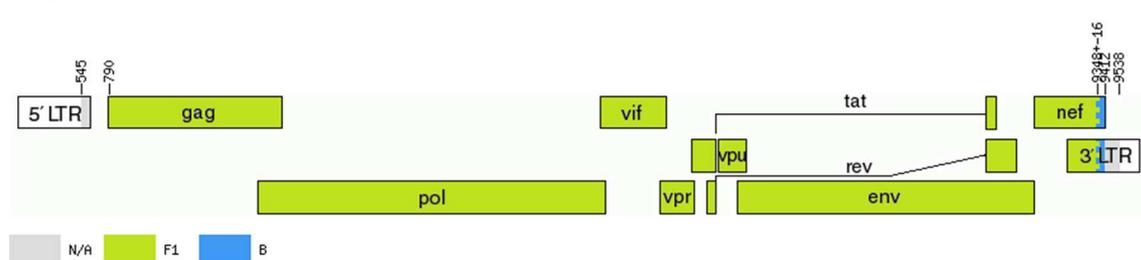


**DQ085873.BR.1999.28**

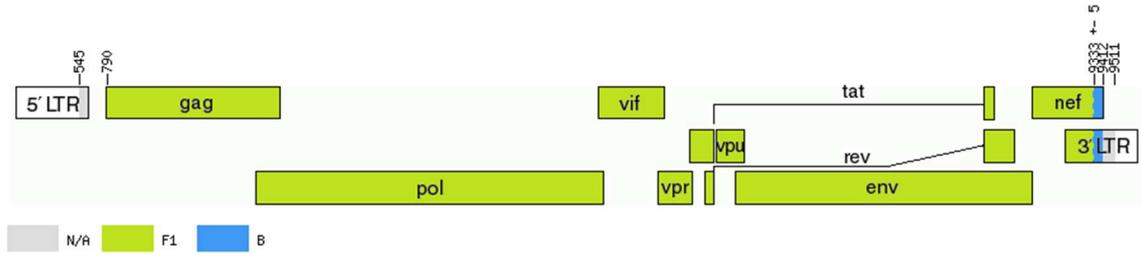


**DQ085874.BR.1999.28**

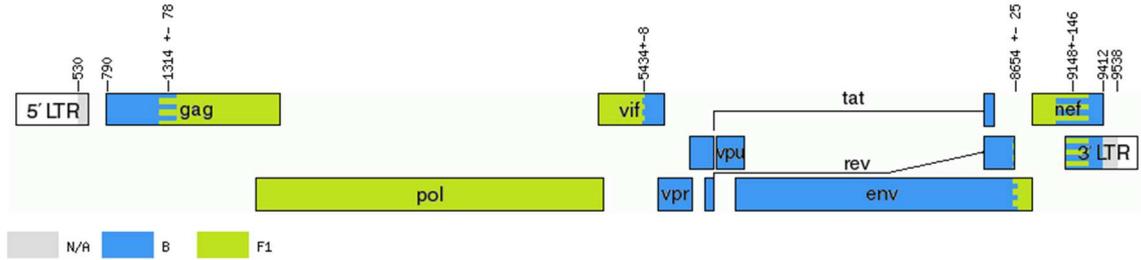


**DQ085875.BR.2000.BF1****DQ085876.BR.2001.29****DQ358799.BR.2001.BF1****DQ358800.BR.2001.BF1****DQ358801.BR.2001.46**

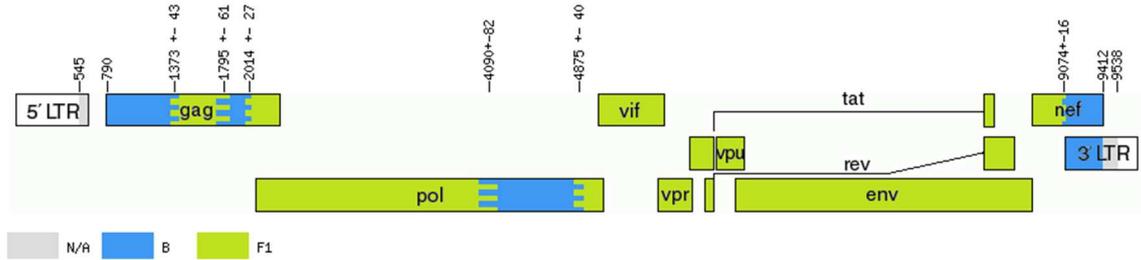
**DQ358802.BR.2001.46**



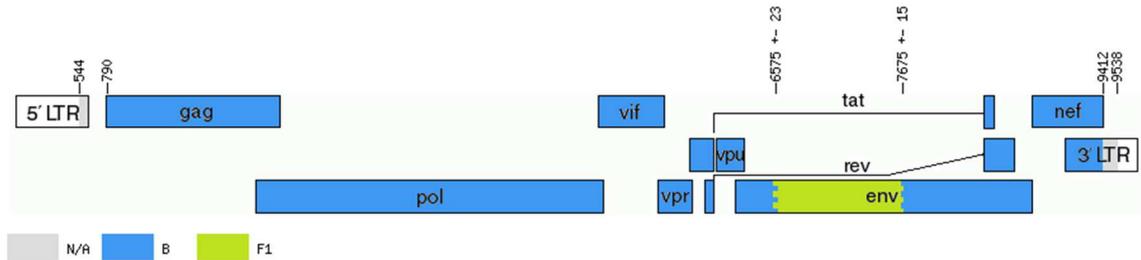
**DQ358803.BR.2001.BF1**



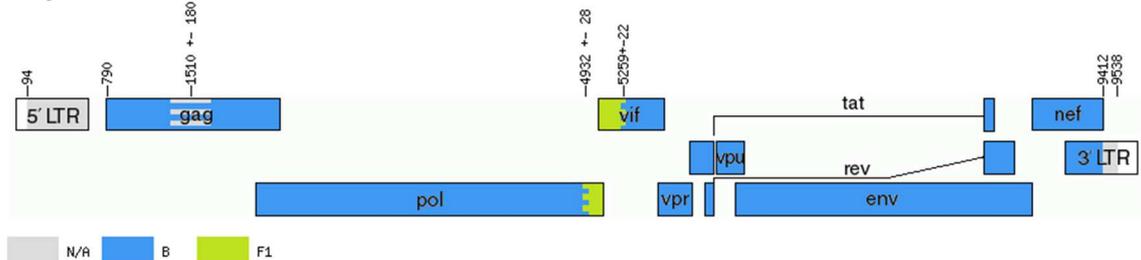
**DQ358804.BR.2001.BF1**



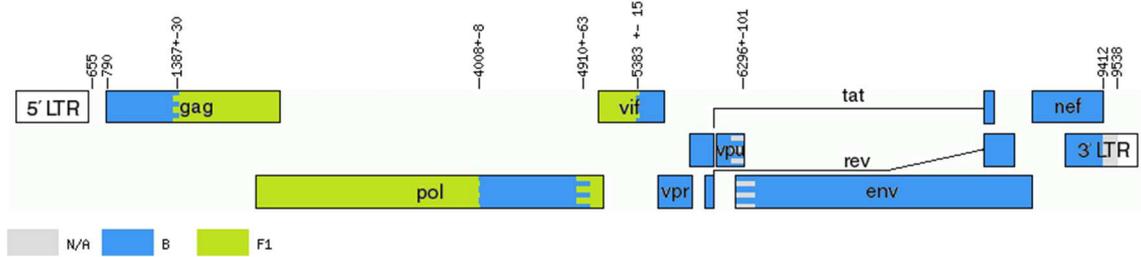
**DQ358806.BR.2002.BF1**



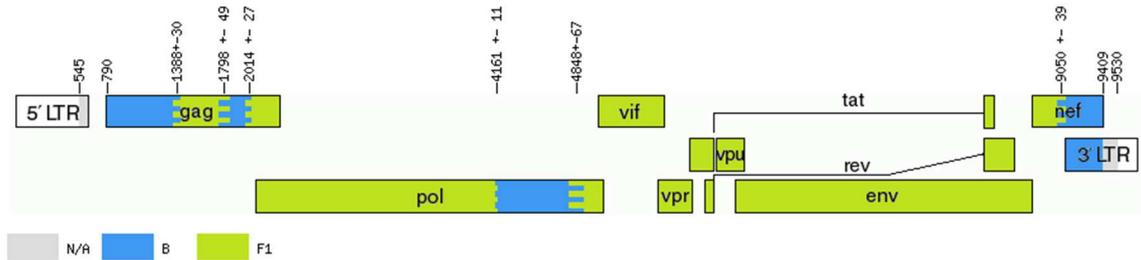
**DQ358807.BR.2002.BF1**



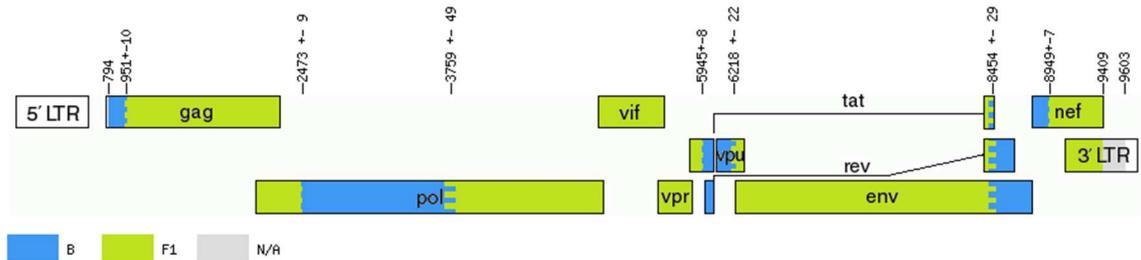
**DQ358811.BR.2002.BF1**



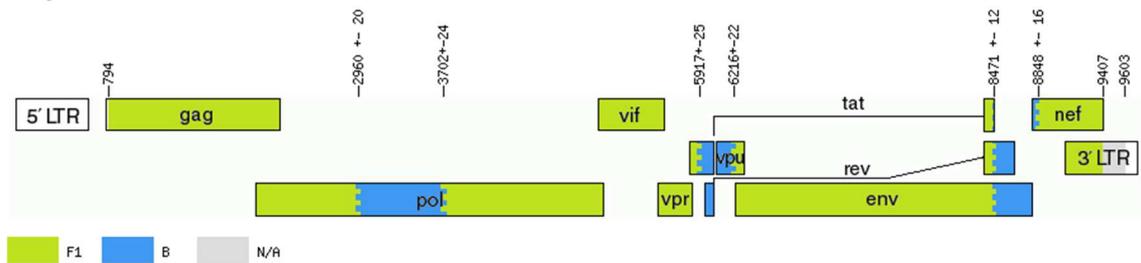
**DQ358812.BR.2002.BF1**



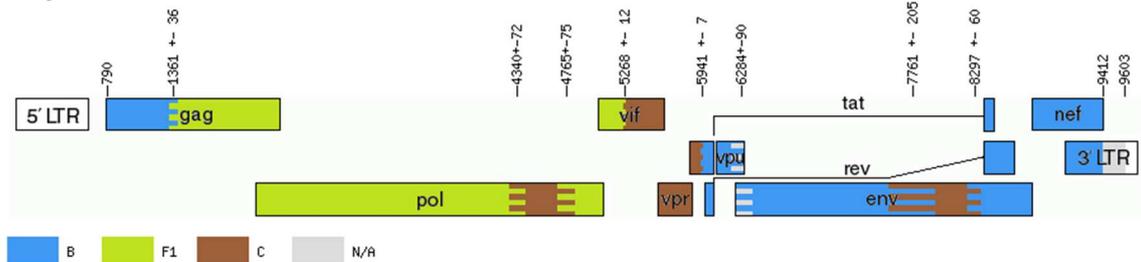
**DQ383747.AR.2002.BF1**

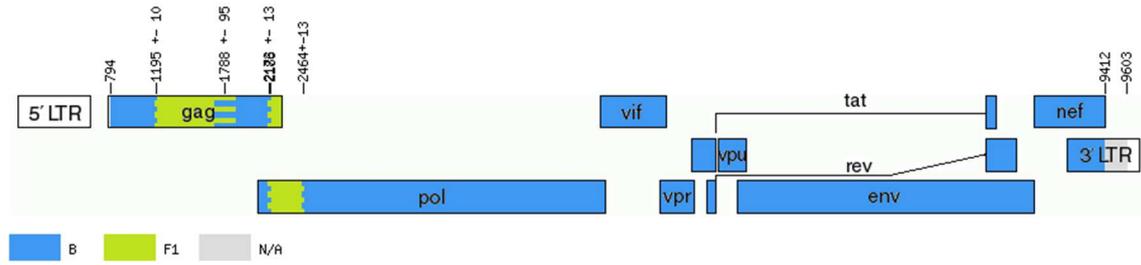
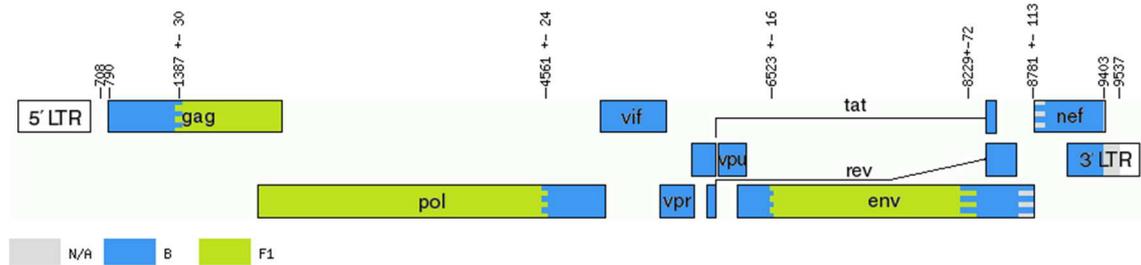
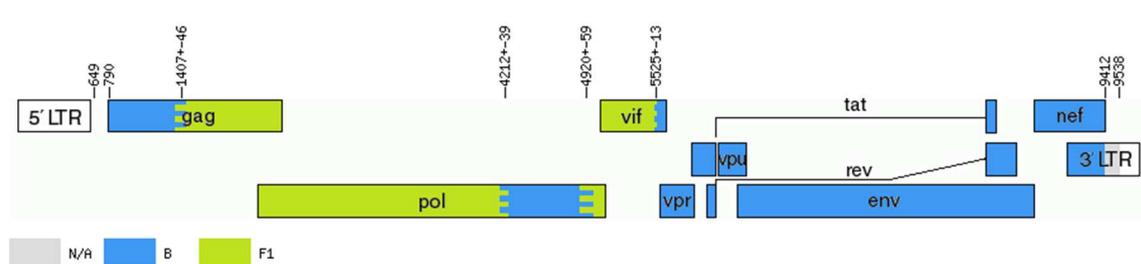
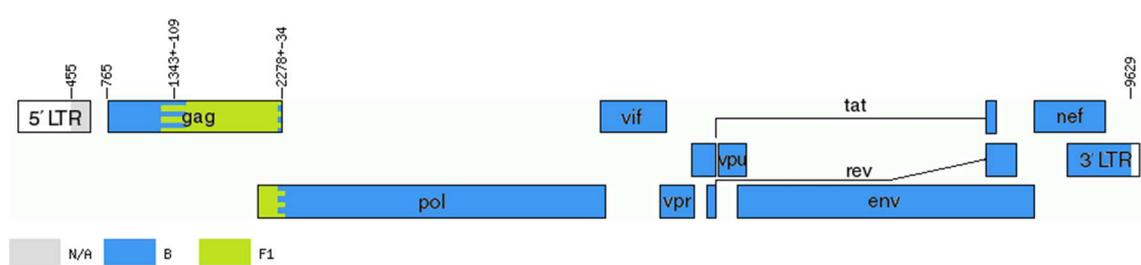
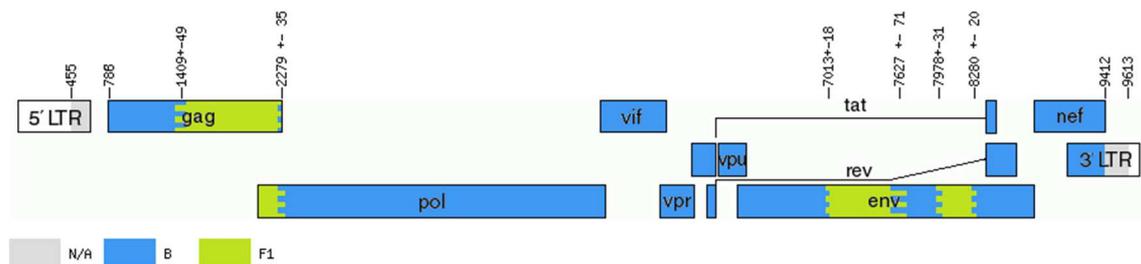


**DQ383753.AR.2004.BF1**

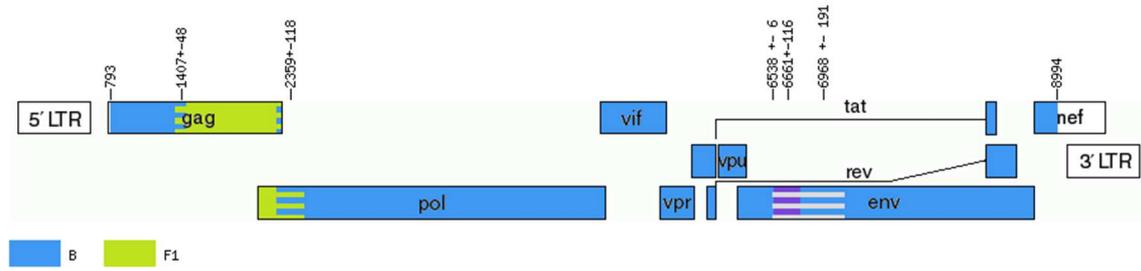


**DQ383754.AR.2004.BCF1**

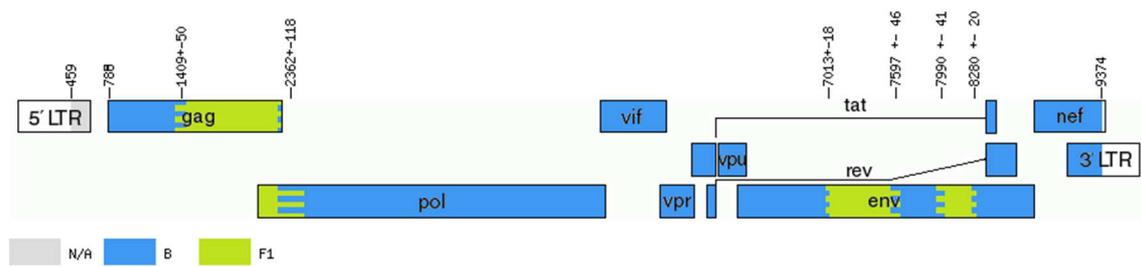


**DQ383755.AR.2005.BF1****EF637052.BR.2003.BF****EF637055.BR.2003.BF****EU170135.LX.2005.42****EU170136.LX.2005.42**

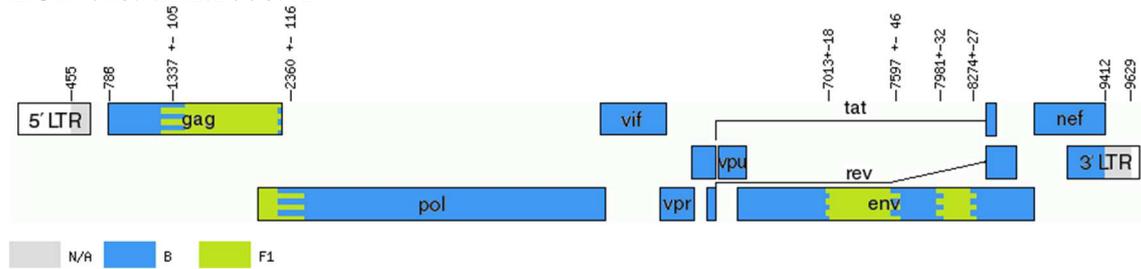
## EU170137.LX.2006.42



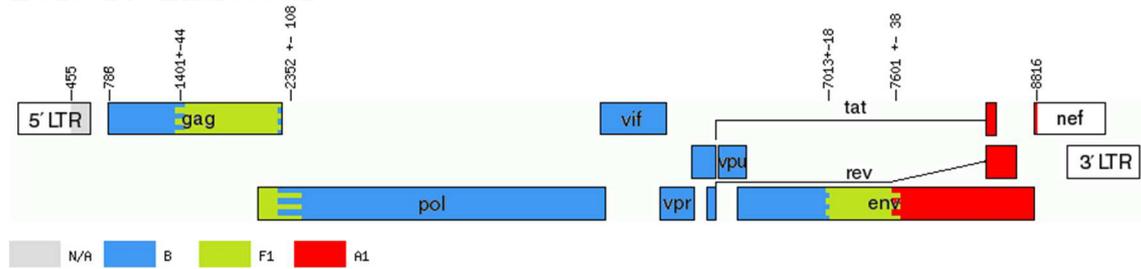
## EU170138.LX.2006.42



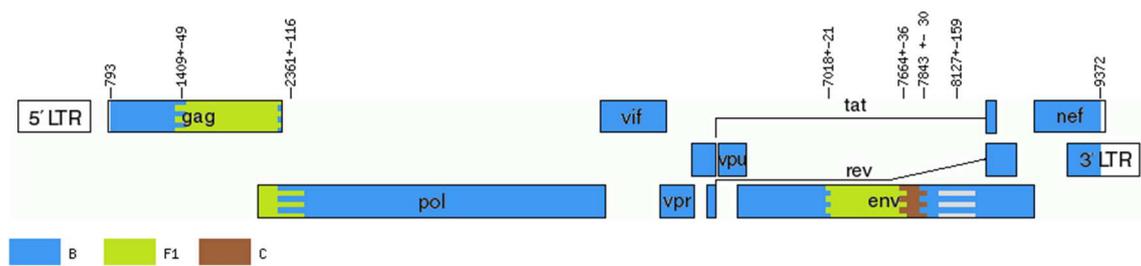
## EU170139.LX.2006.42



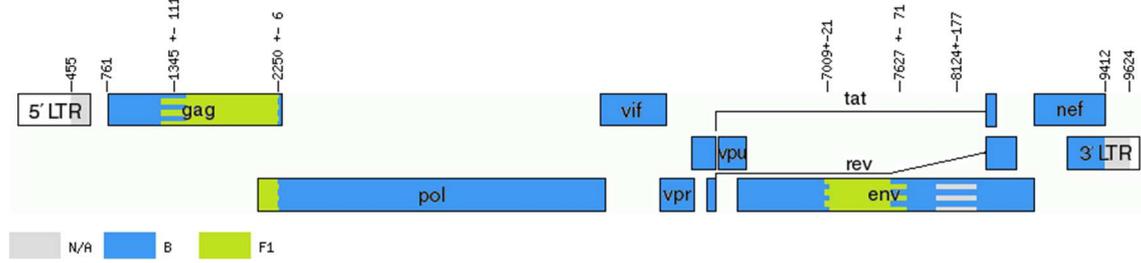
## EU170140.LX.2003.42



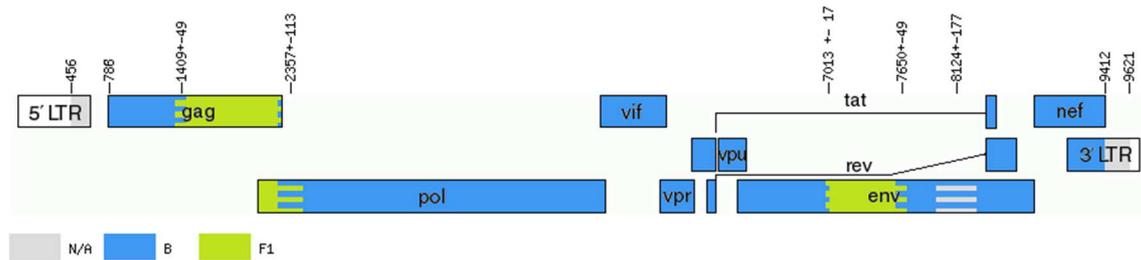
## EU170141.LX.2006.42



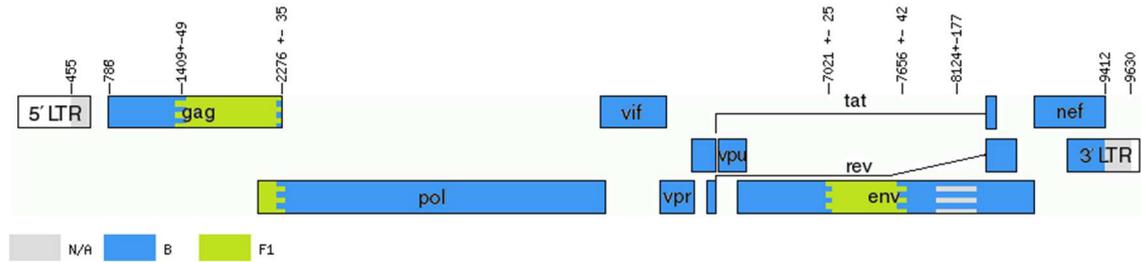
## EU170142.LX.2004.42



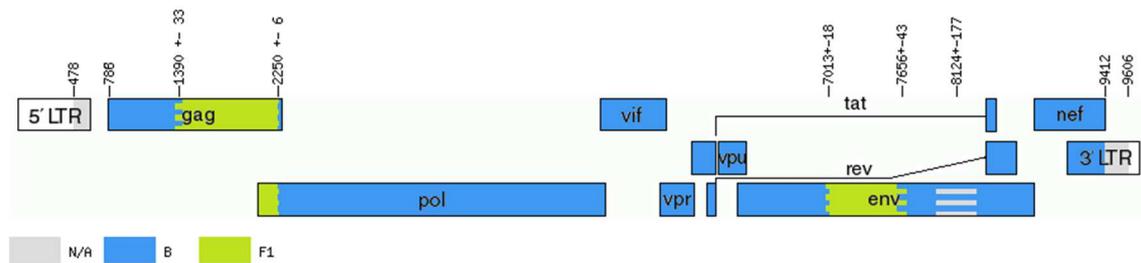
## EU170143.LX.2005.42



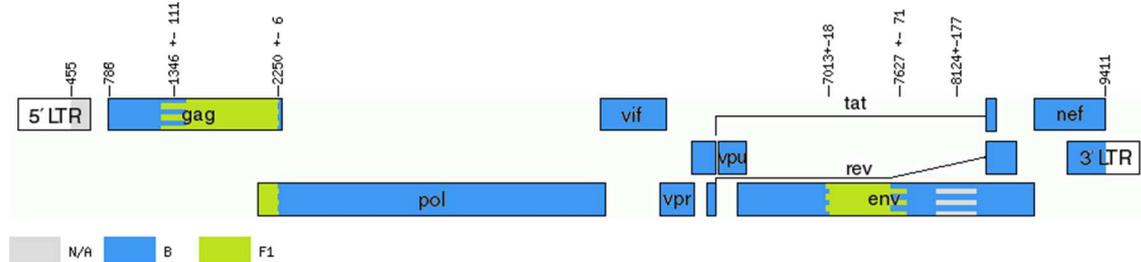
## EU170144.LX.2005.42



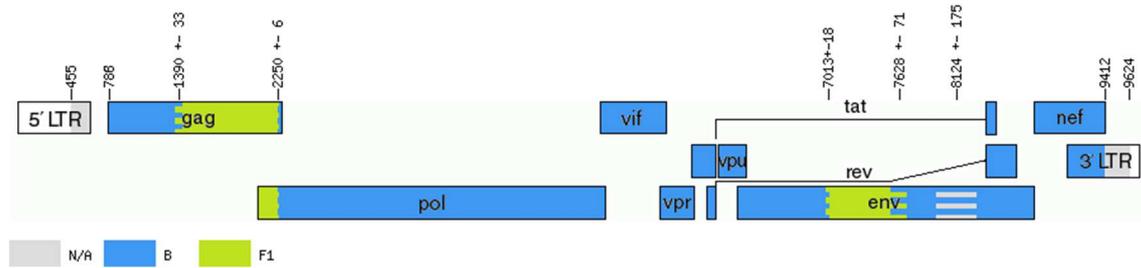
## EU170145.LX.2003.42



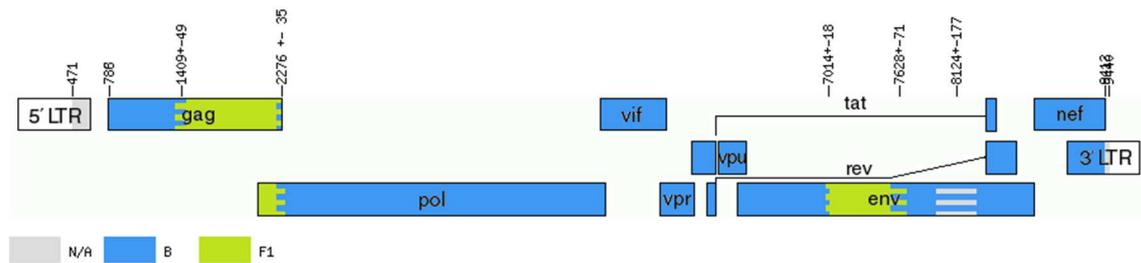
## EU170146.LX.2004.42



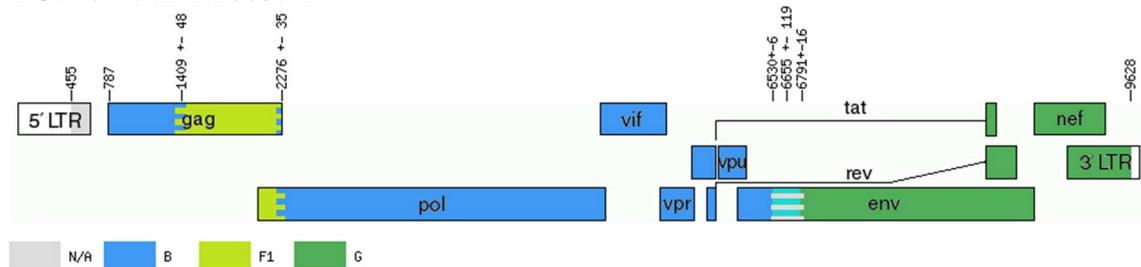
## EU170147.LX.2003.42



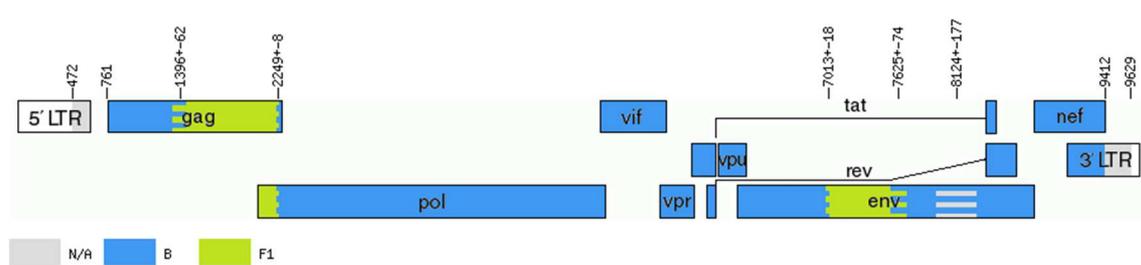
## EU170148.LX.2003.42



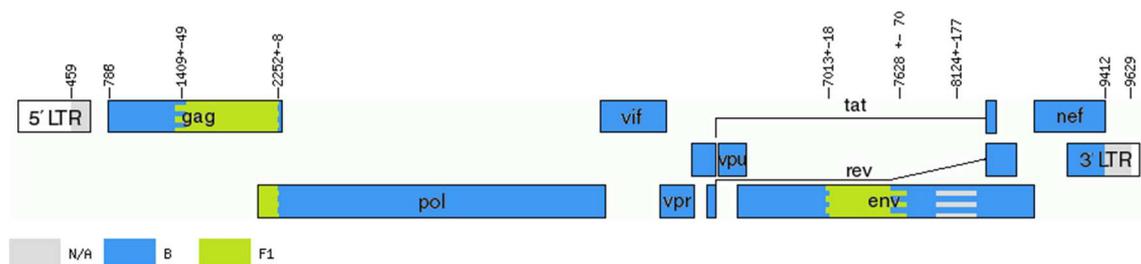
## EU170149.LX.2003.42



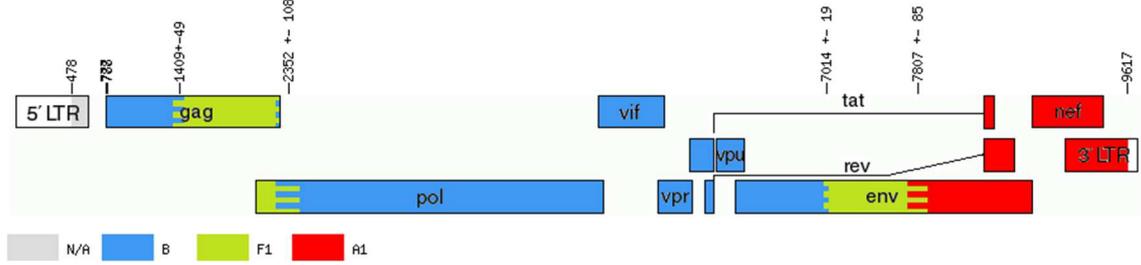
## EU170150.LX.2003.42



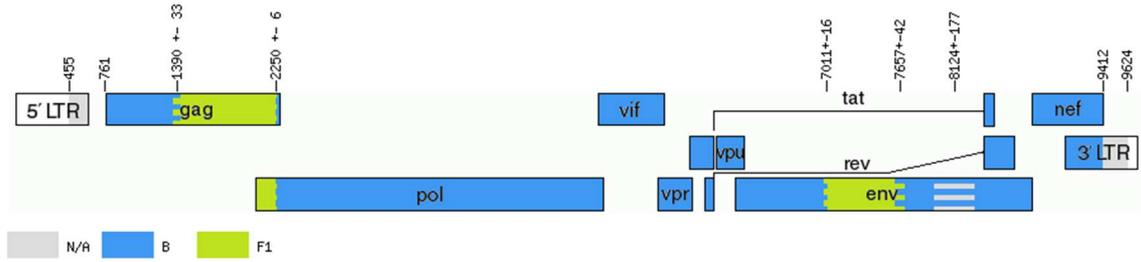
## EU170151.LX.2003.42



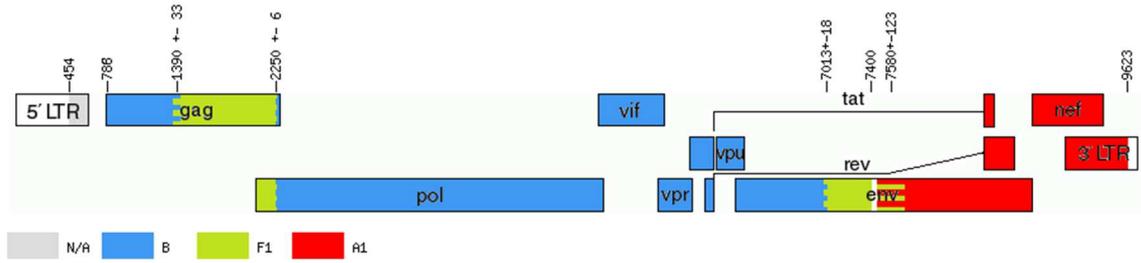
**EU170152.LX.2003.42**



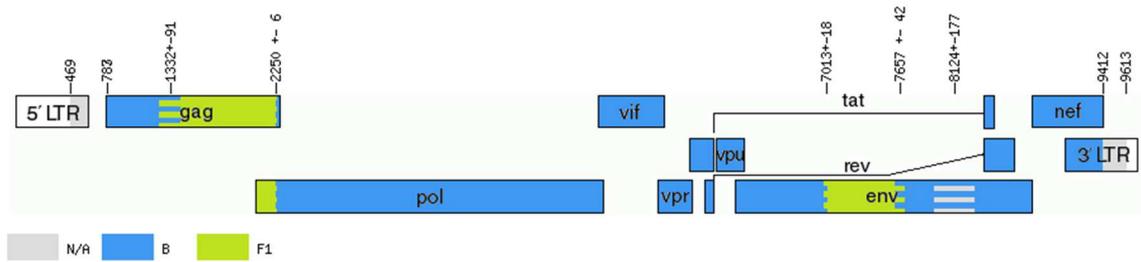
**EU170153.LX.2005.42**



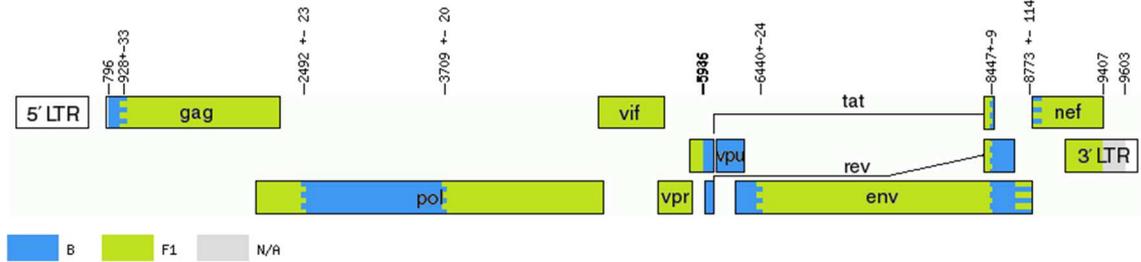
**EU170154.LX.2003.42**



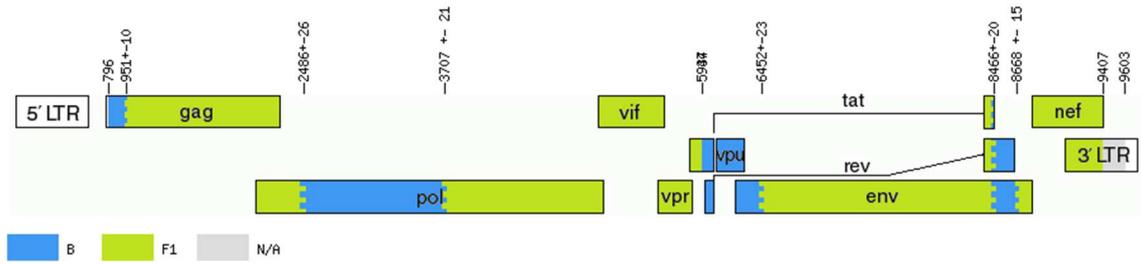
**EU170155.LX.2003.42**



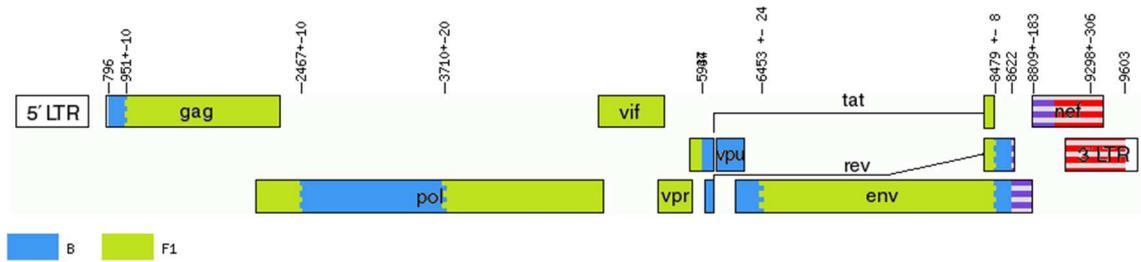
**EU581823.PR.2002.17**



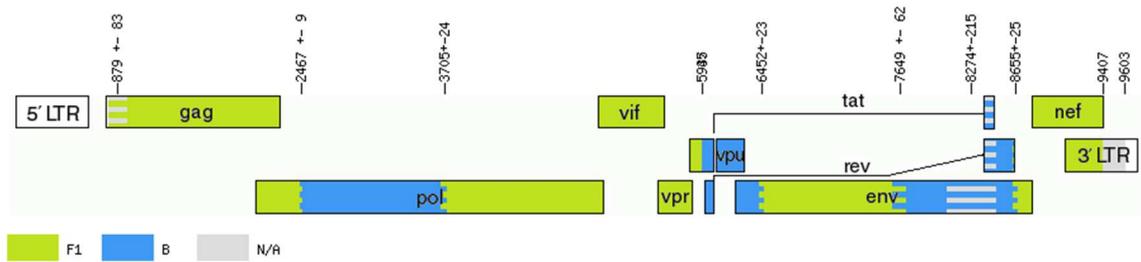
**EU581824.PR.2002.17**



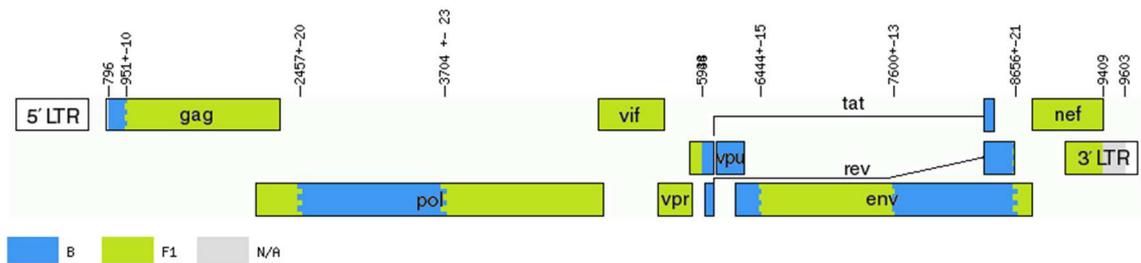
**EU581825.AR.2002.17**



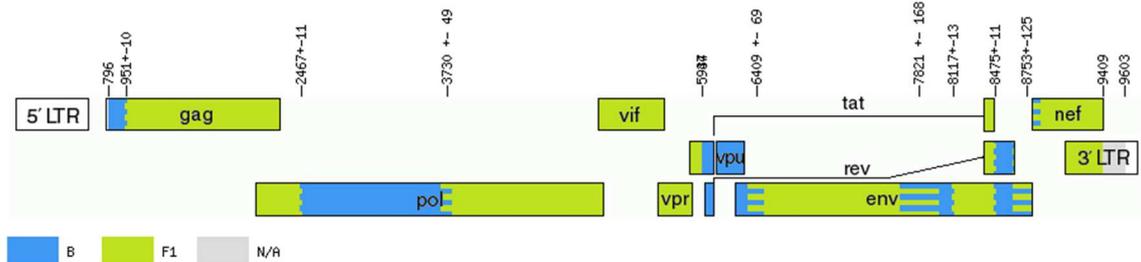
**EU581826.AR.2002.17**



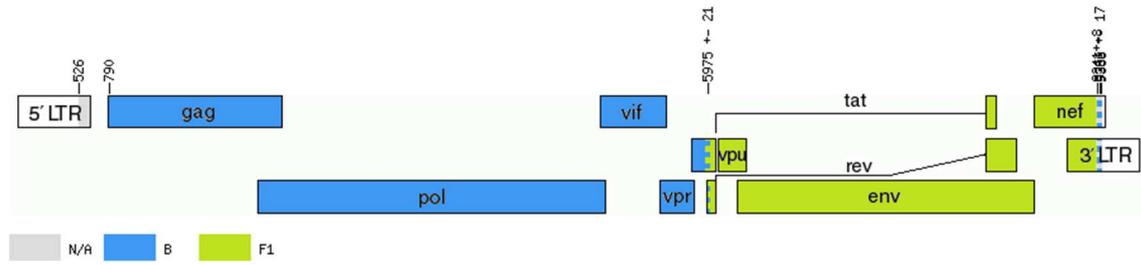
**EU581827.BO.2002.17**



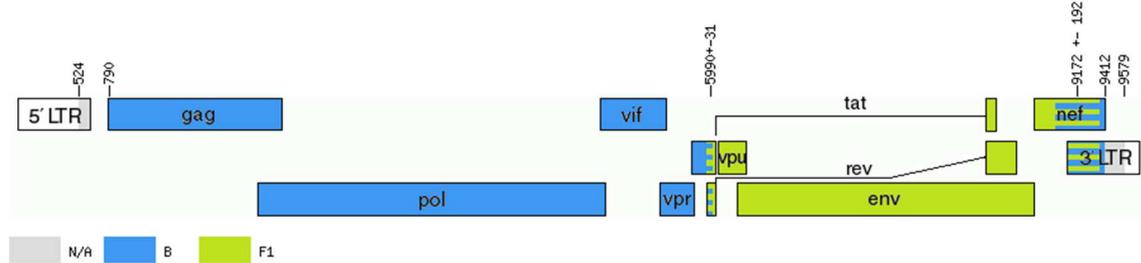
**EU581828.PE.2002.17**



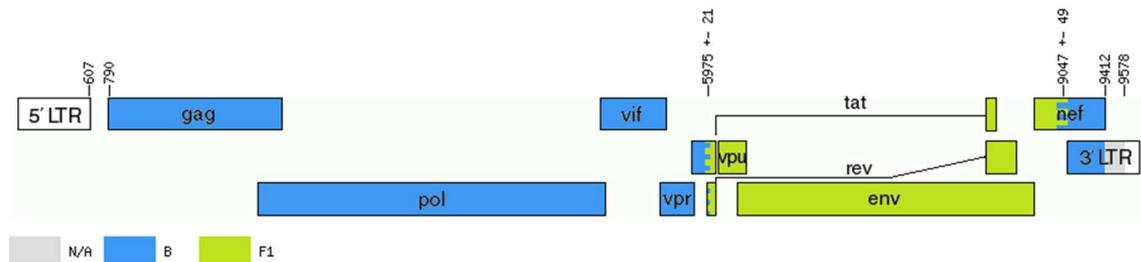
## EU735534.BR.2003.39



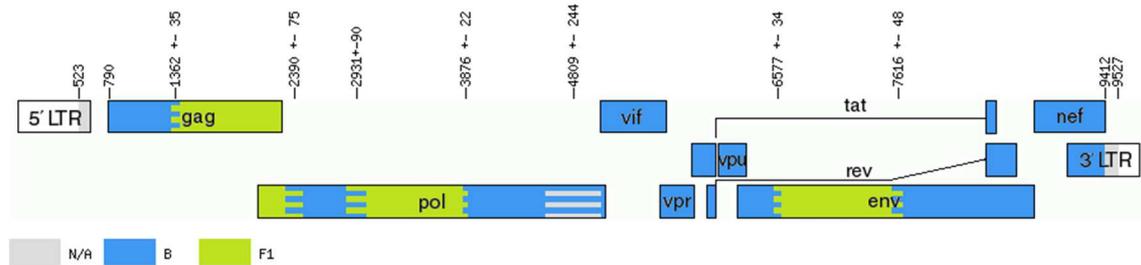
## EU735535.BR.2004.39



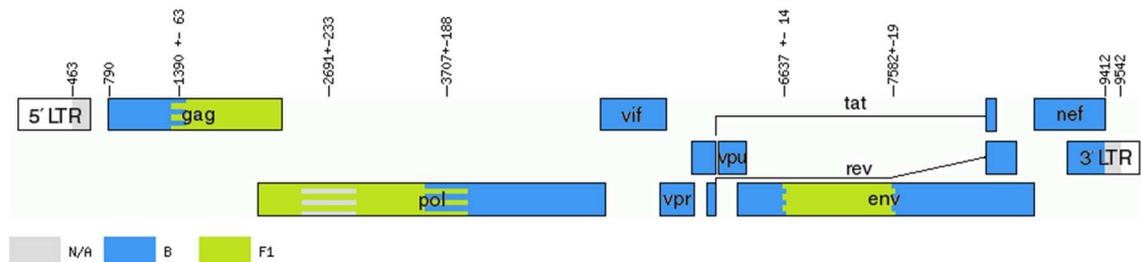
## EU735536.BR.2003.39



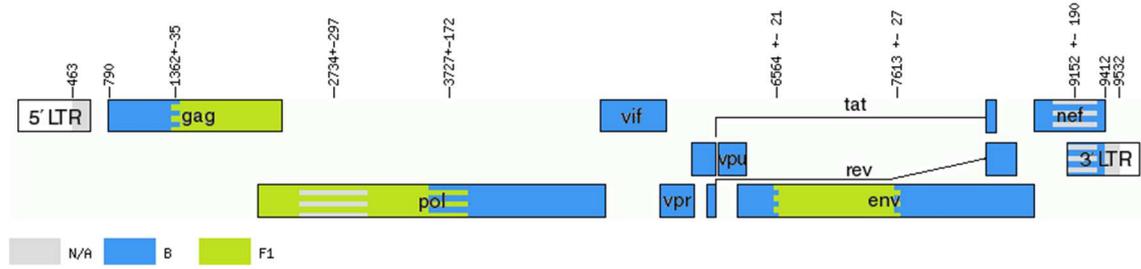
## EU735537.BR.2005.40



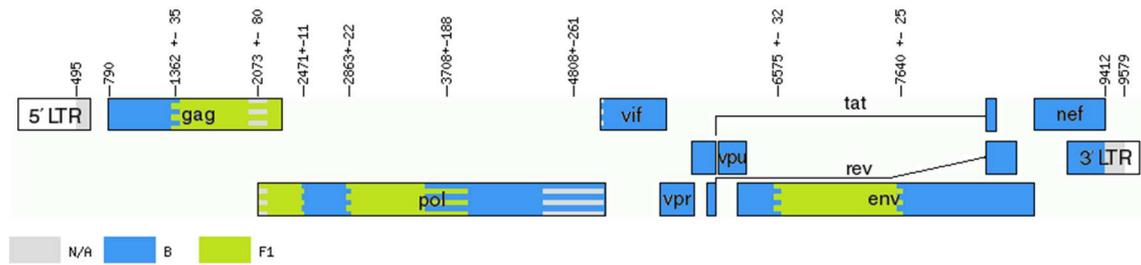
## EU735538.BR.2004.40



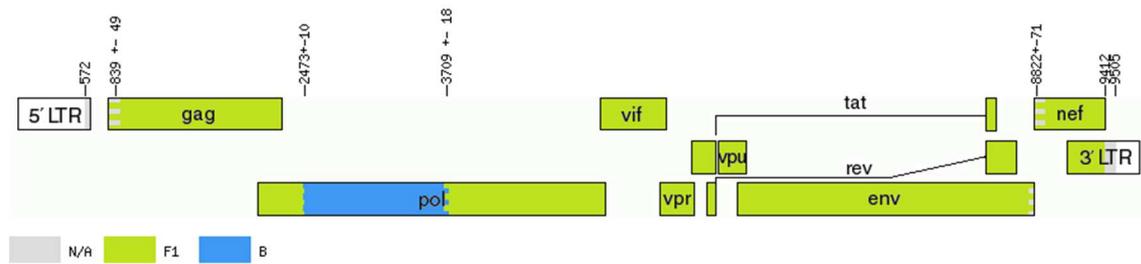
## EU735539.BR.2005.40



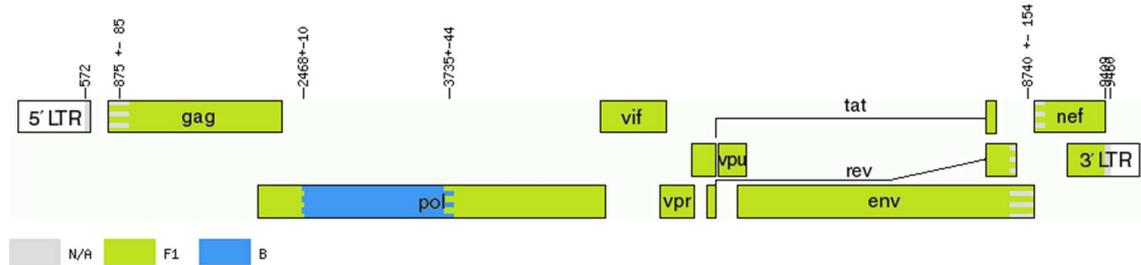
## EU735540.BR.2004.40



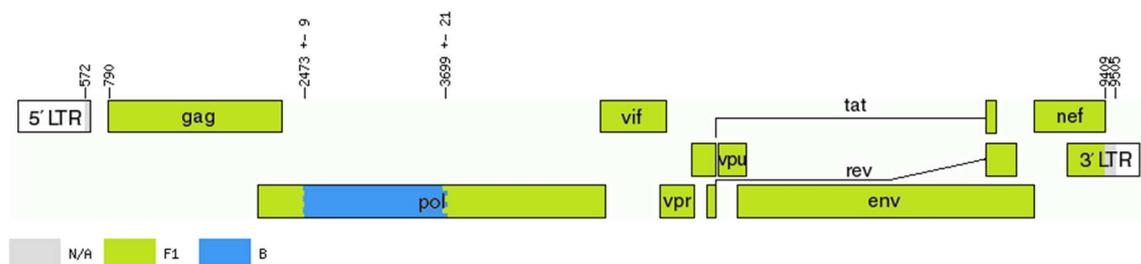
## FJ213780.UR.2005.38

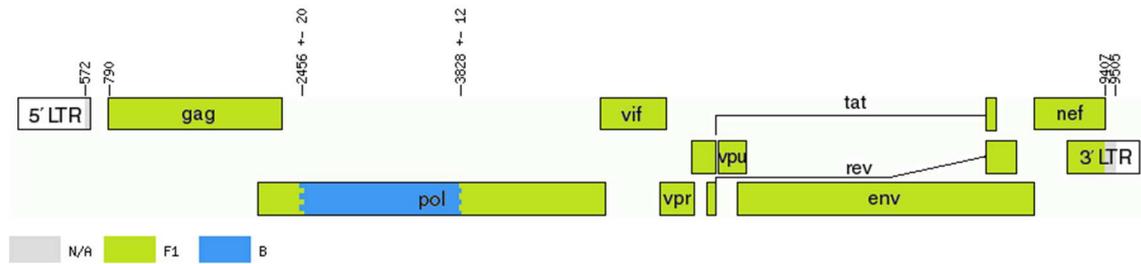
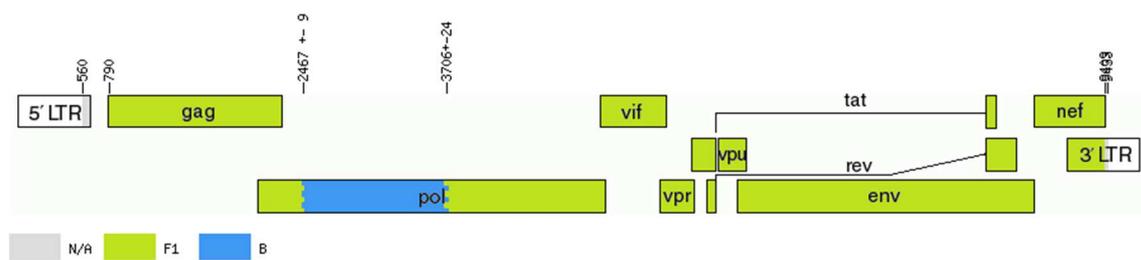
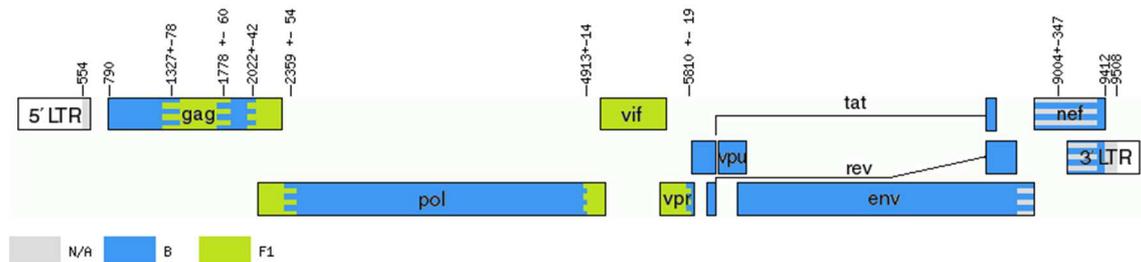
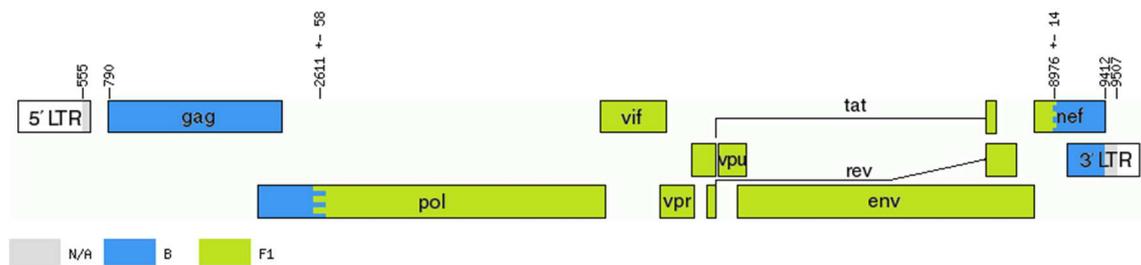
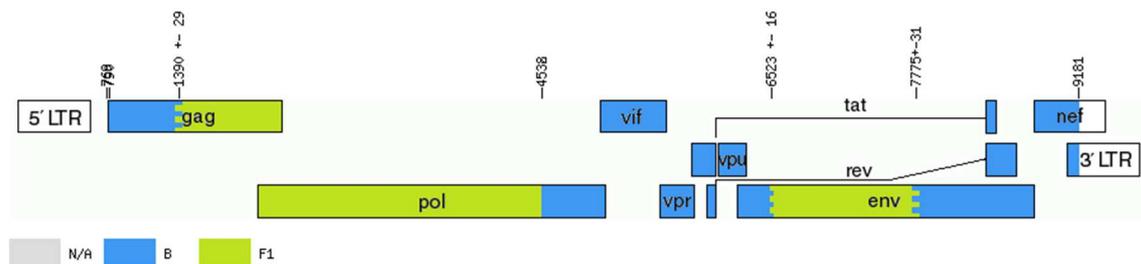


## FJ213781.UR.2004.38

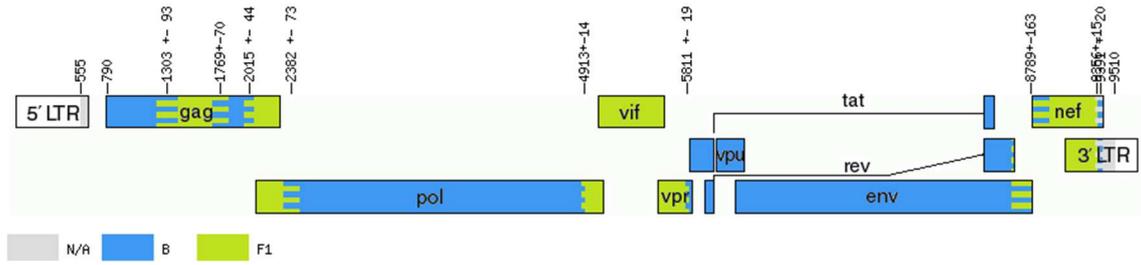


## FJ213782.UR.2004.38

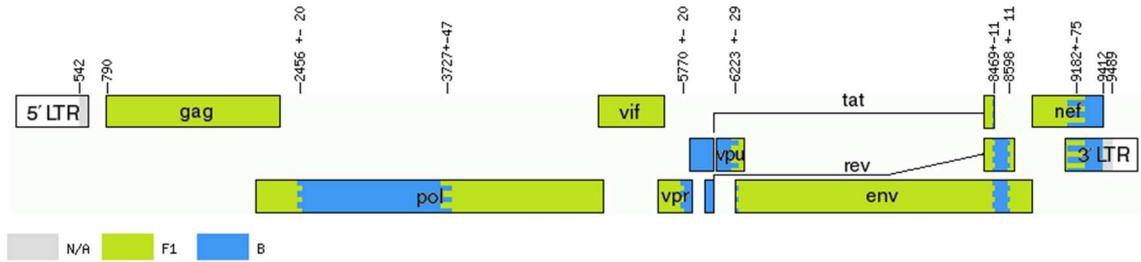


**FJ213783.UR.2003.38****FJ358521.CH.2000.44****FJ670529.ES.2008.47****FJ853621.ES.2008.BF1****FJ904244.IT.2008.BF1**

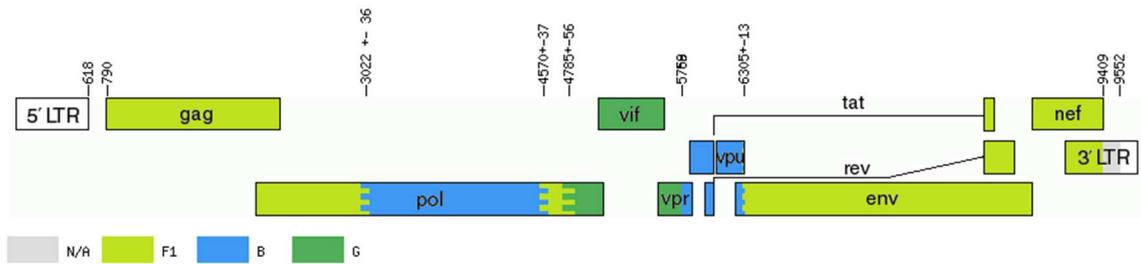
**GQ372987.ES.2008.47**



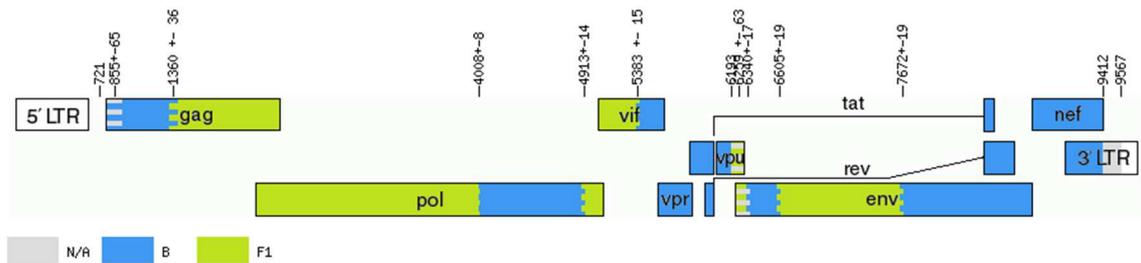
**GQ372989.ES.2008.BF1**



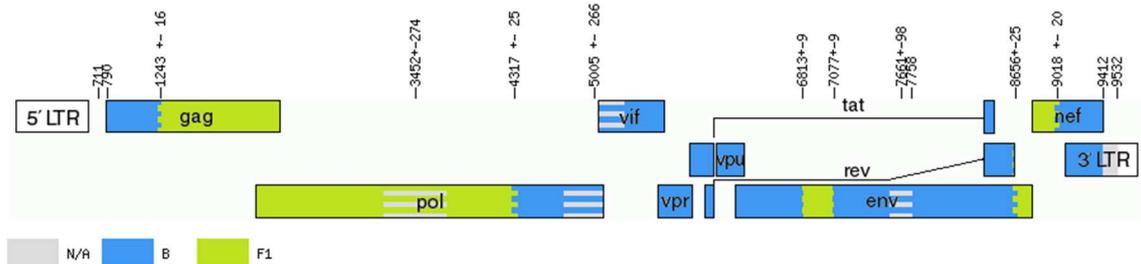
**GU207884.MC.2005.BF1G**



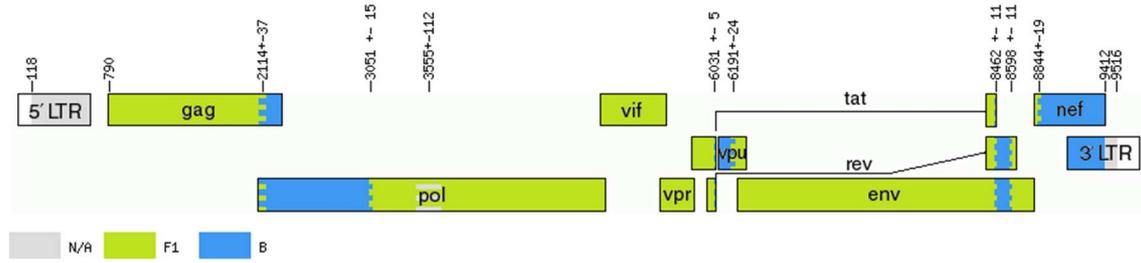
**GU595148.IT.2002.BF1**



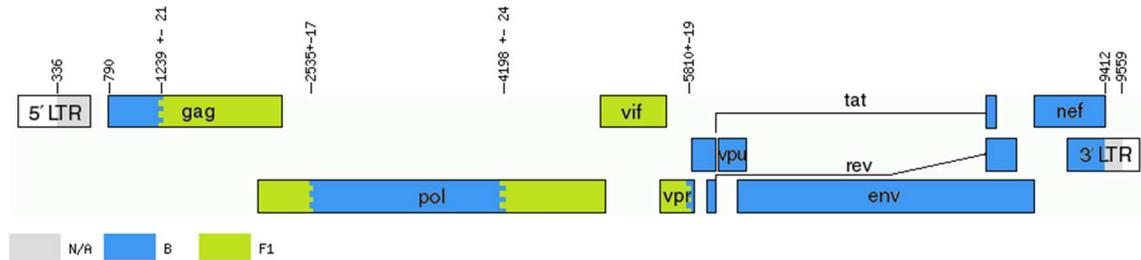
**GU595149.IT.2001.BF1**



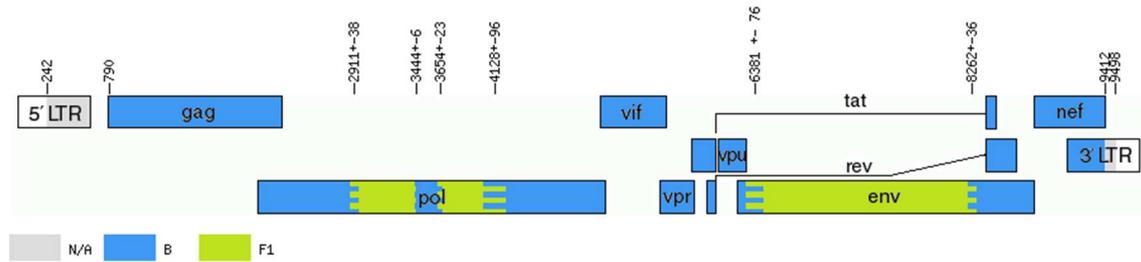
### GU595150.IT.2002.BF1



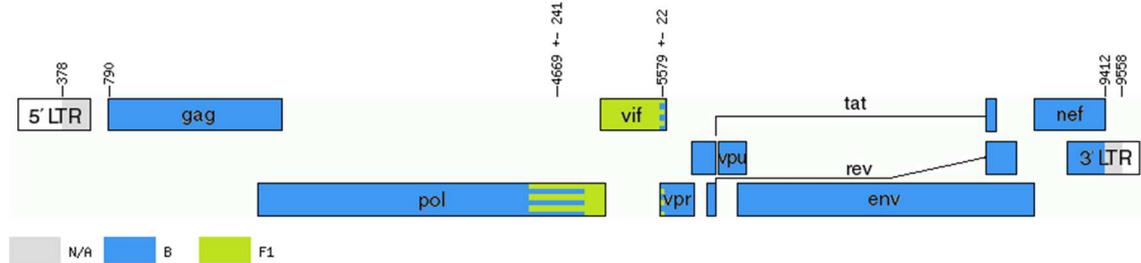
### GU595151.IT.2002.29



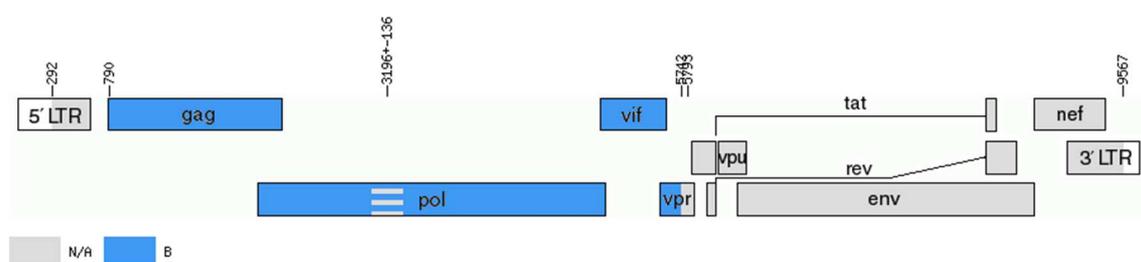
### GU595152.IT.2005.39

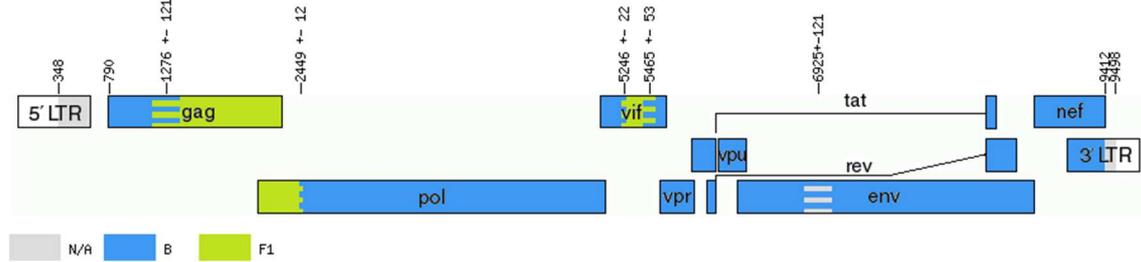
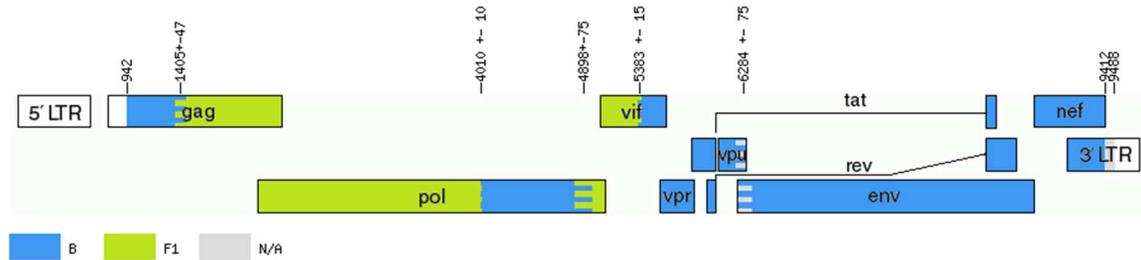
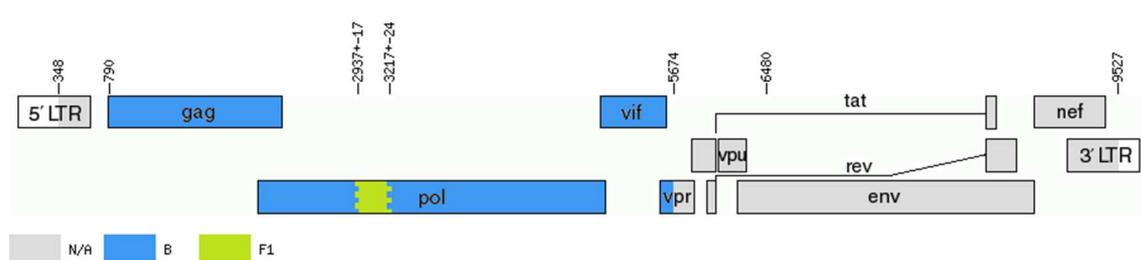
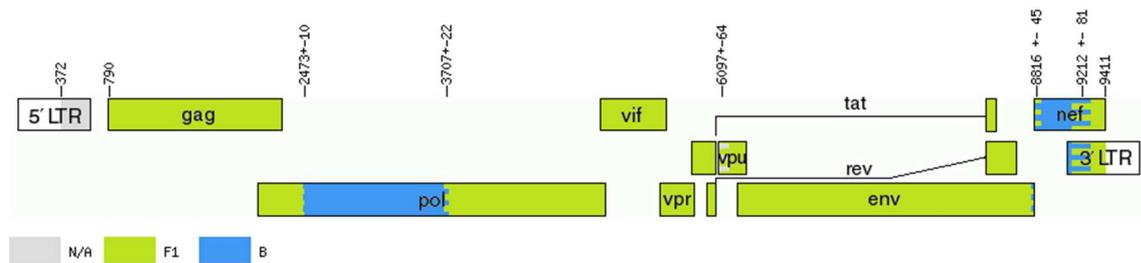
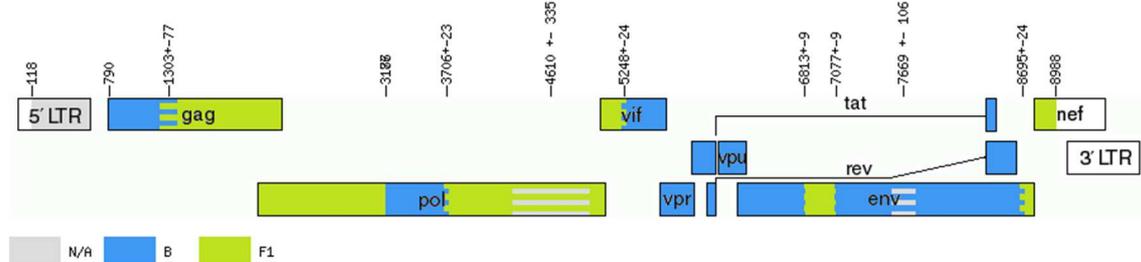


### GU595153.IT.2006.BF1

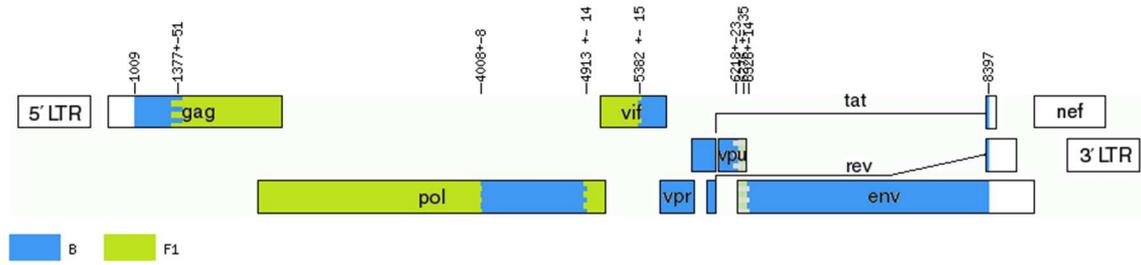


### GU595154.IT.2002.BF

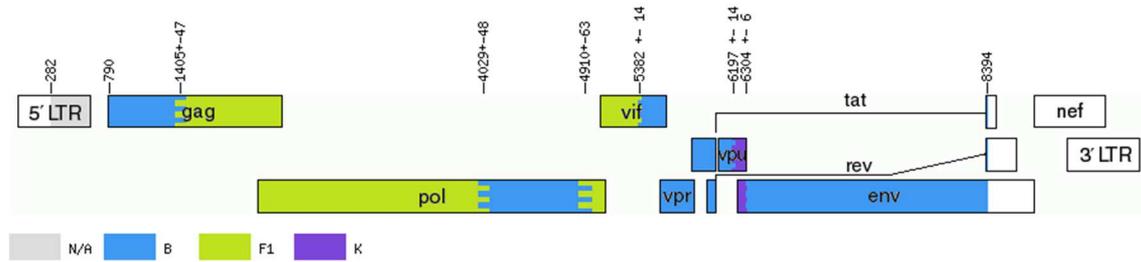


**GU595155.IT.2005.BF1****GU595156.IT.2001.BF1****GU595158.IT.2005.BF1****GU595159.IT.2007.BF1****GU595160.IT.2002.BF1**

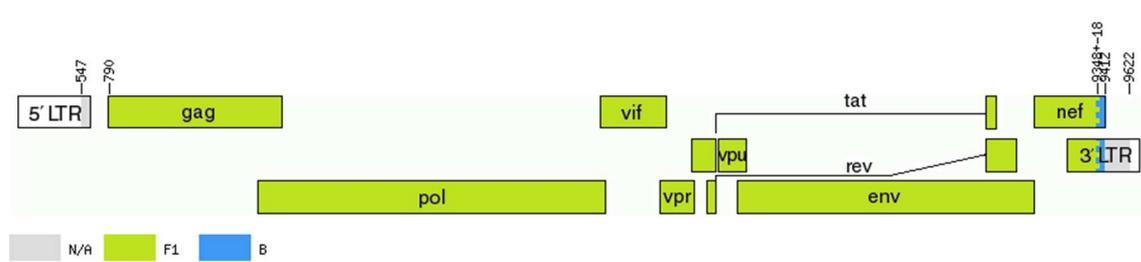
### GU595161.IT.2002.BF1



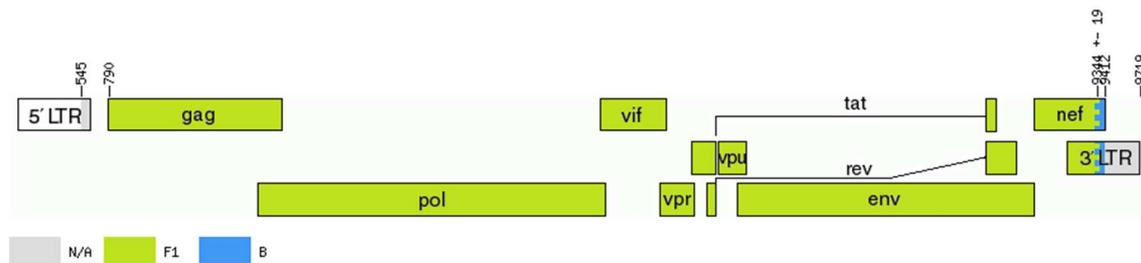
### GU595162.IT.2001.BF1



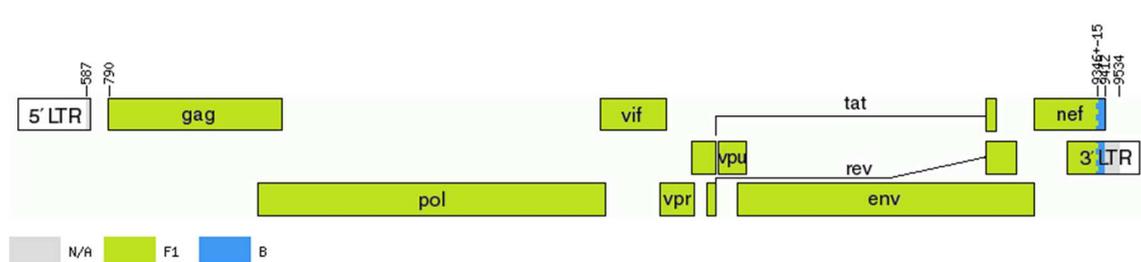
### HM026455.BR.2007.BF1



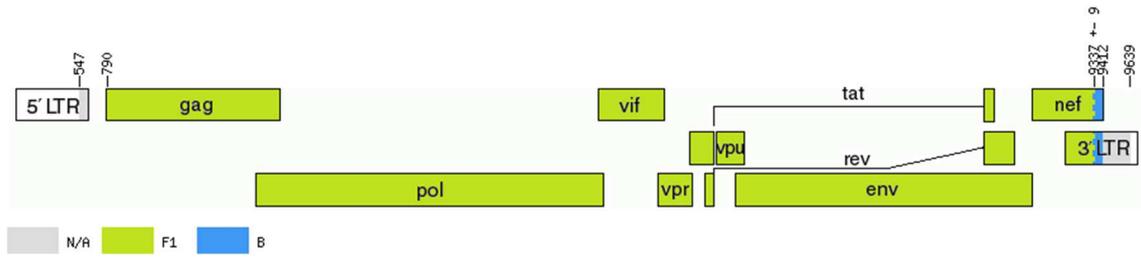
### HM026456.BR.2007.46



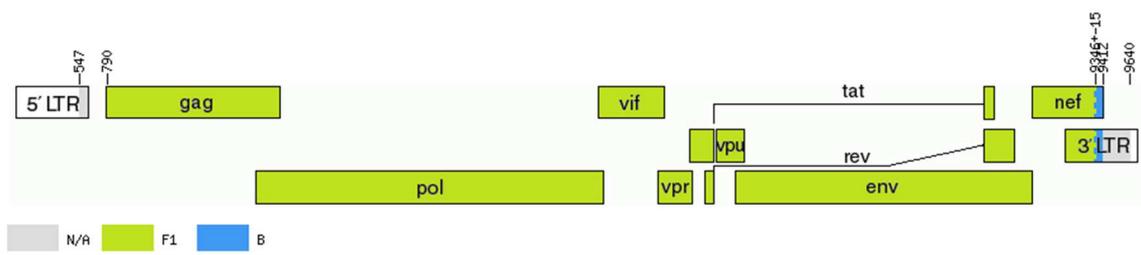
### HM026457.BR.2007.46



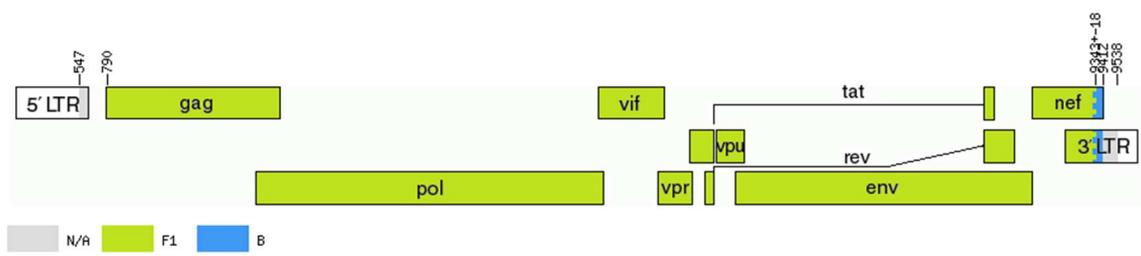
**HM026458.BR.2007.46**



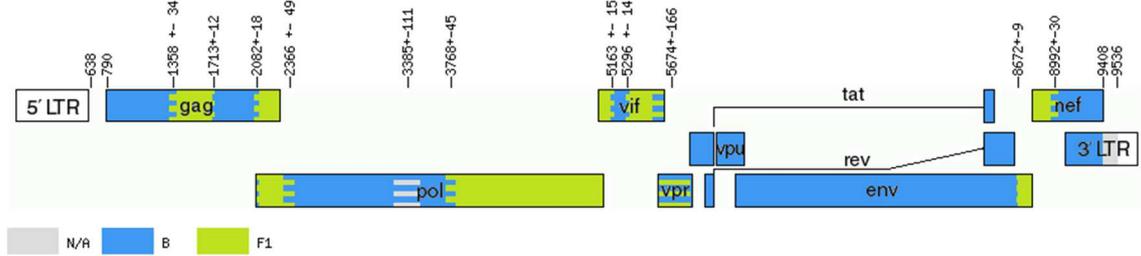
**HM026459.BR.2007.46**



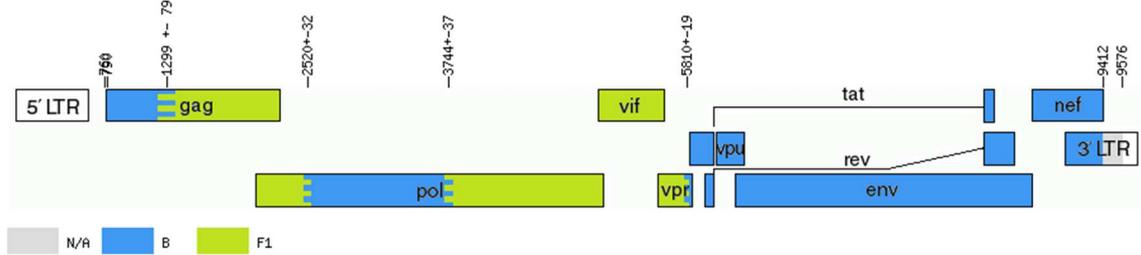
**HM026460.BR.2007.46**

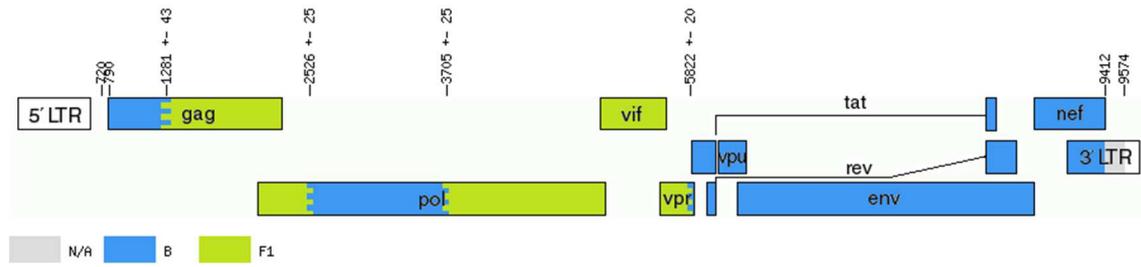
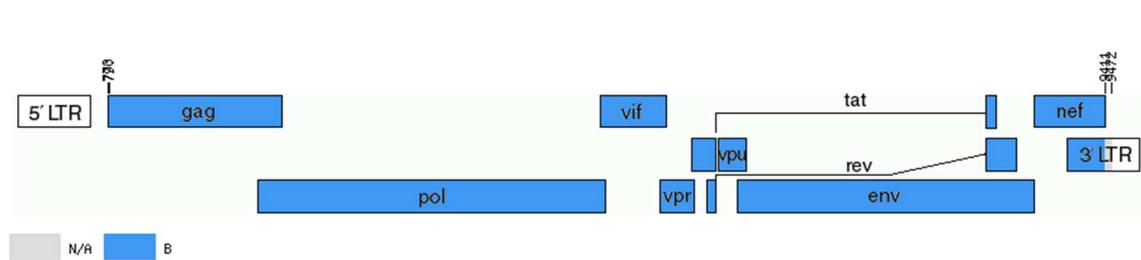
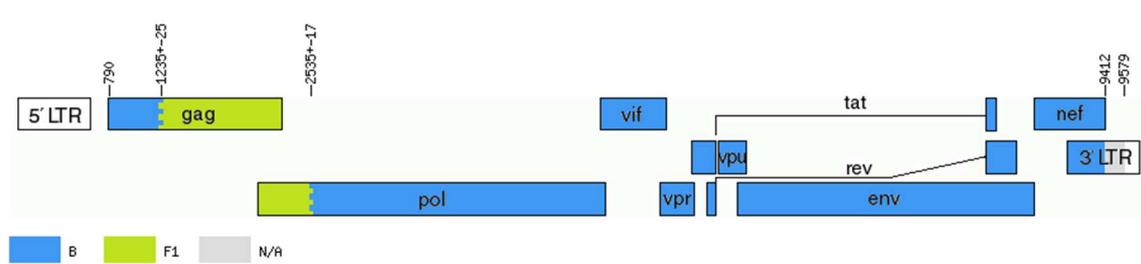
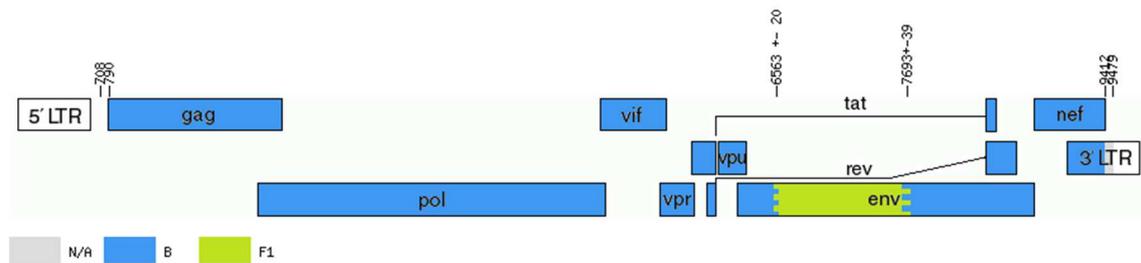
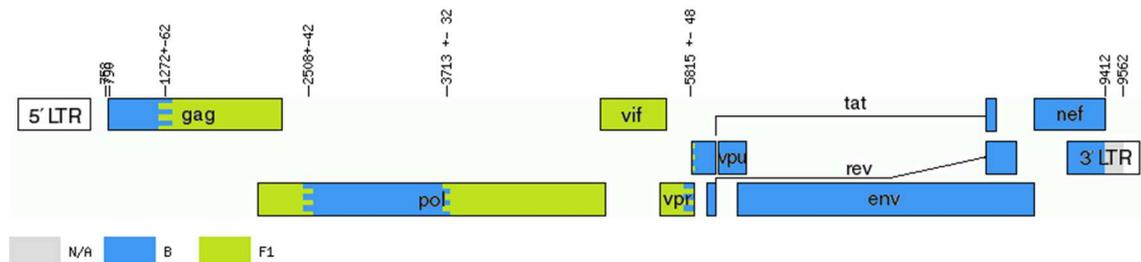


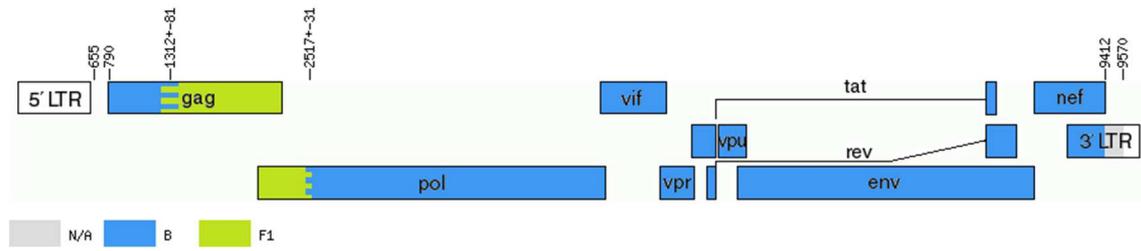
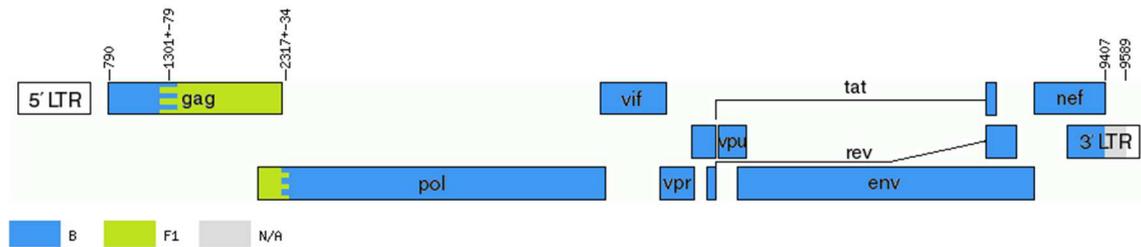
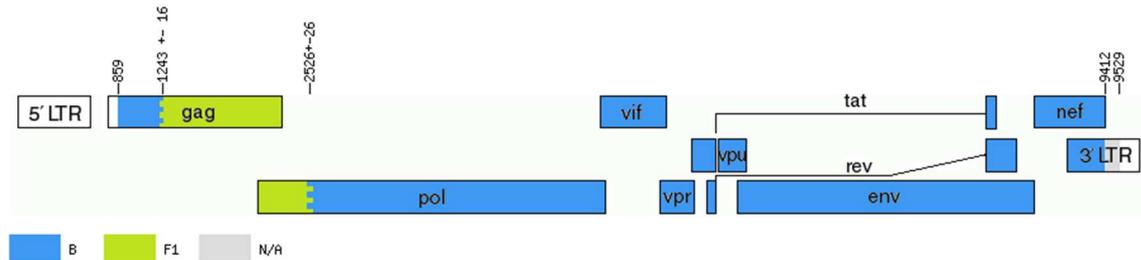
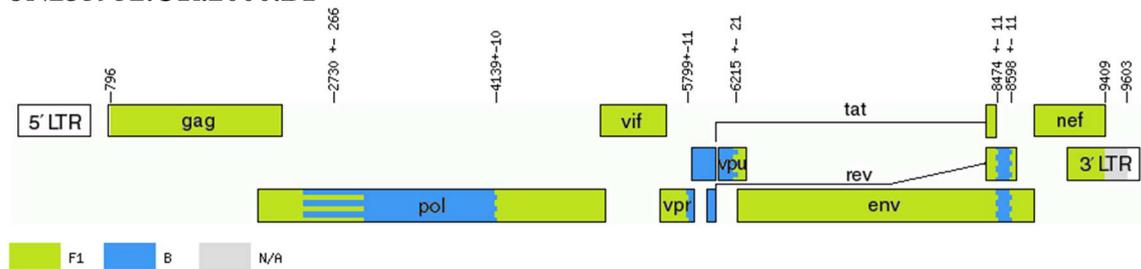
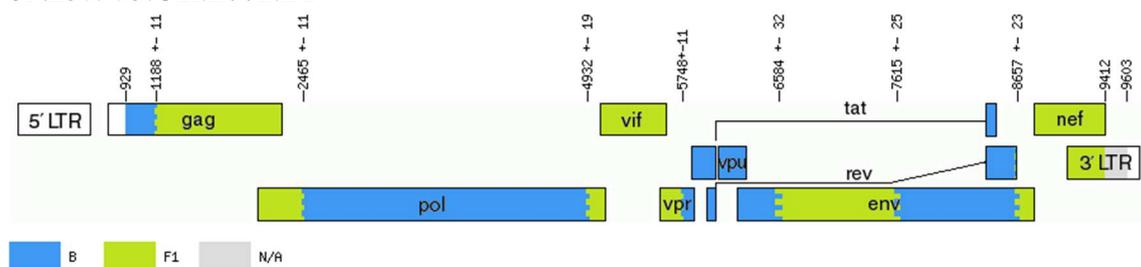
**JF804805.BR.2005.BF1**

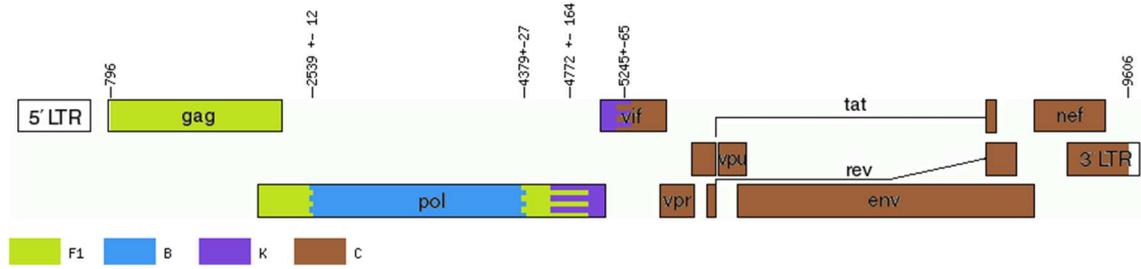
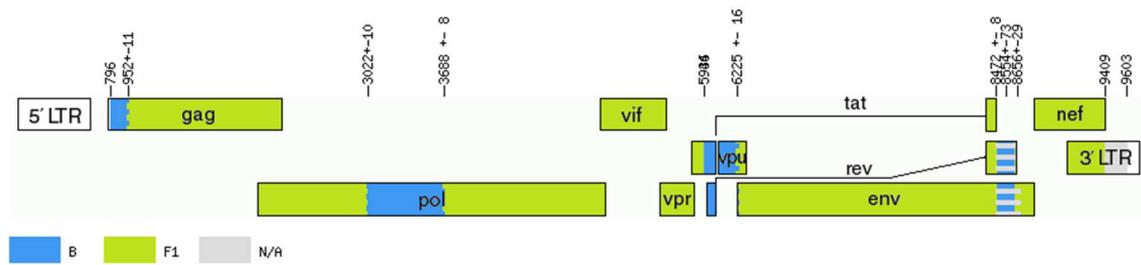
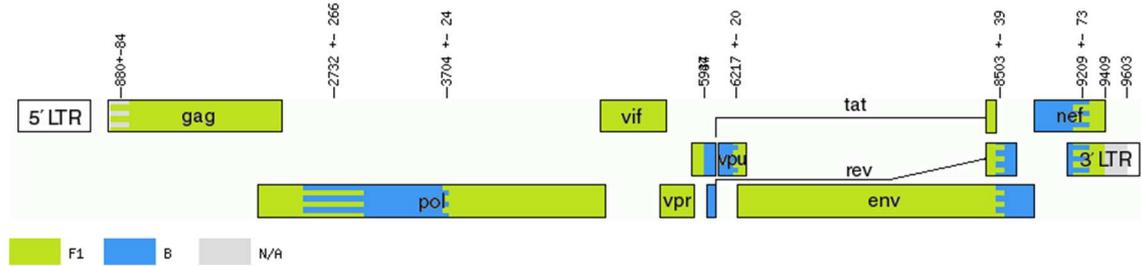
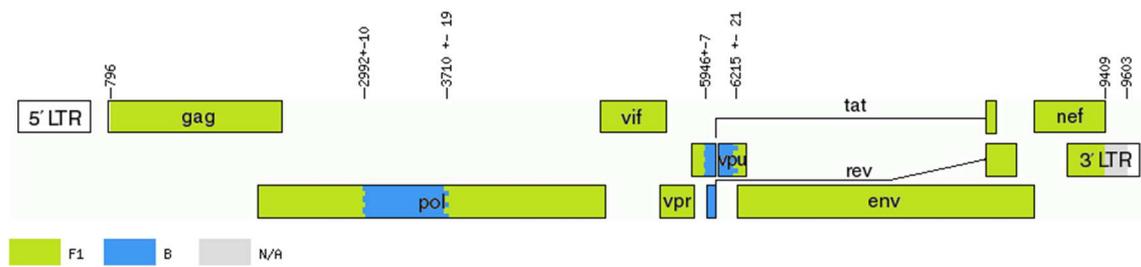
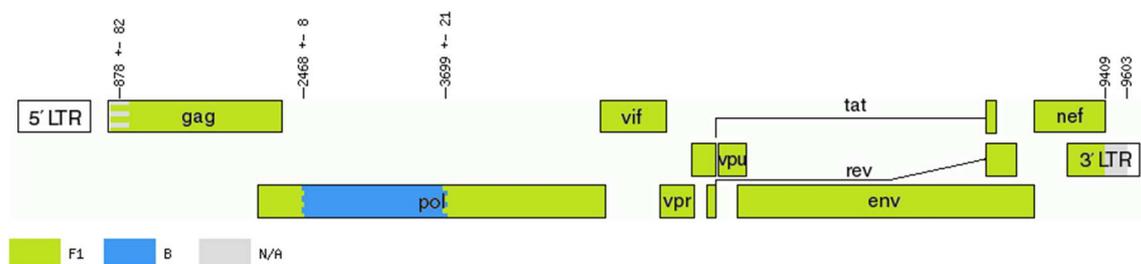


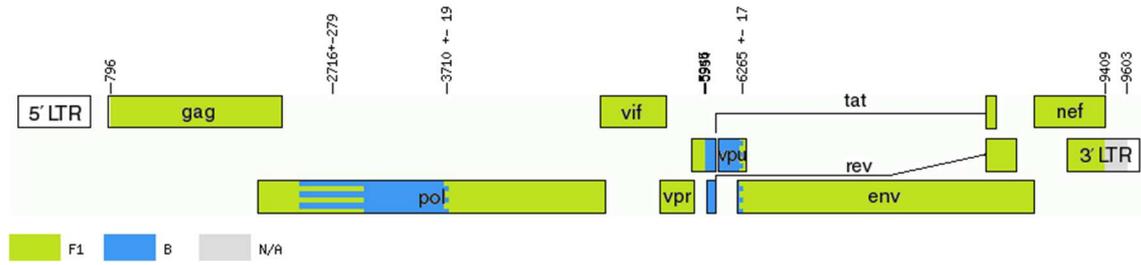
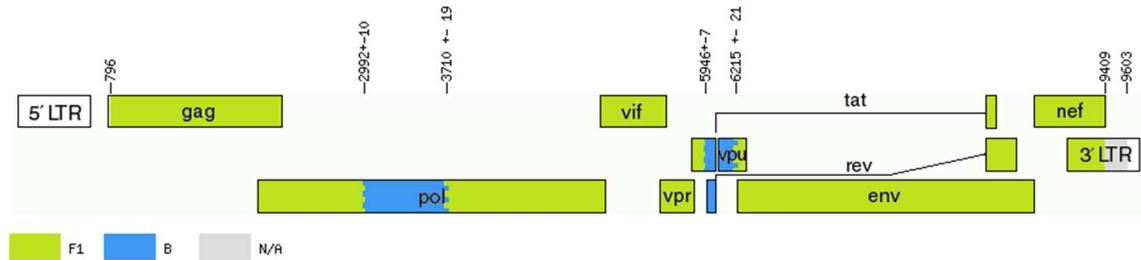
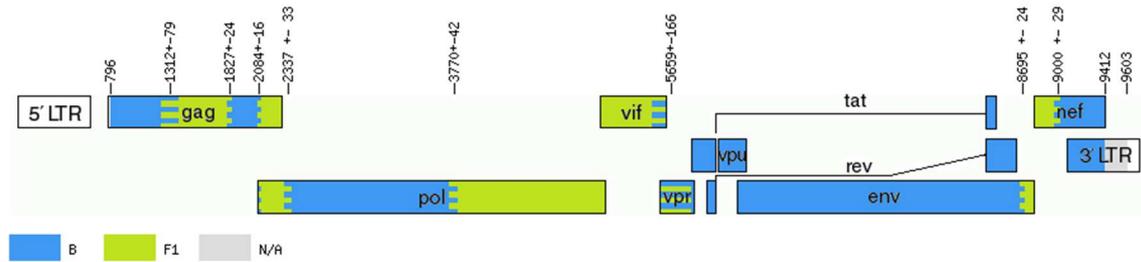
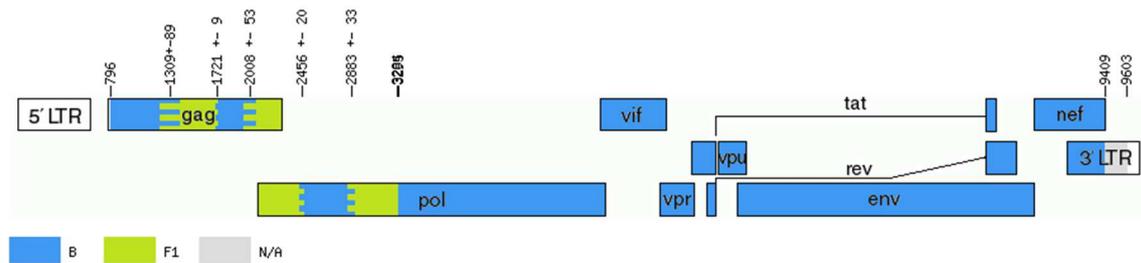
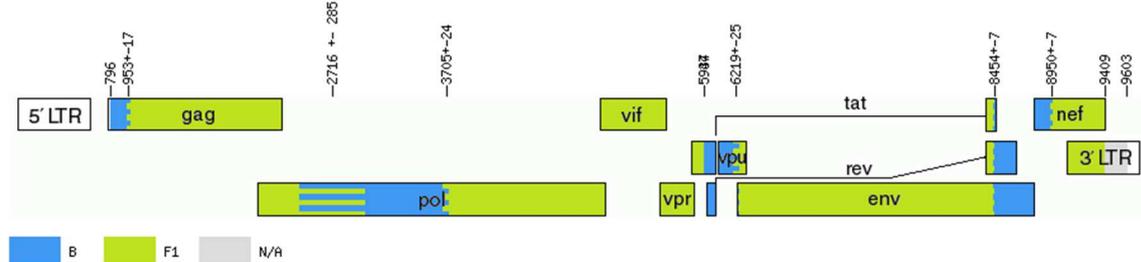
**JF804806.BR.2005.29**

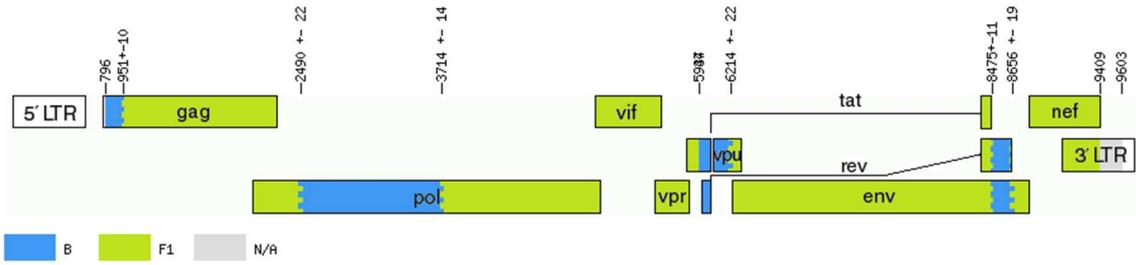
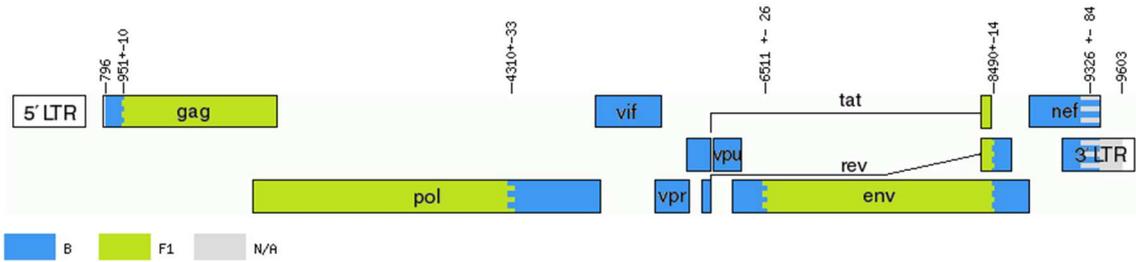
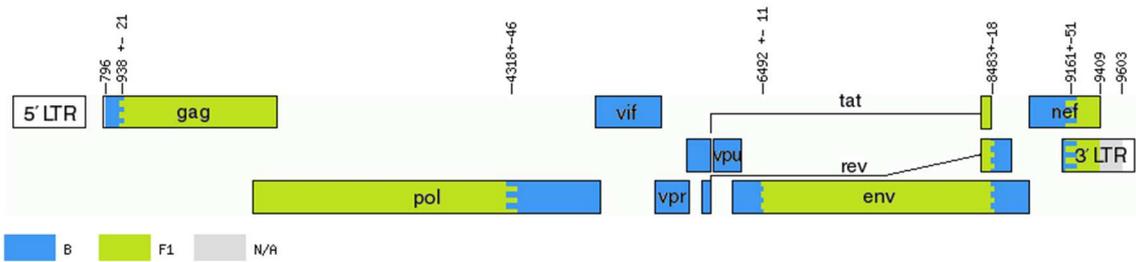
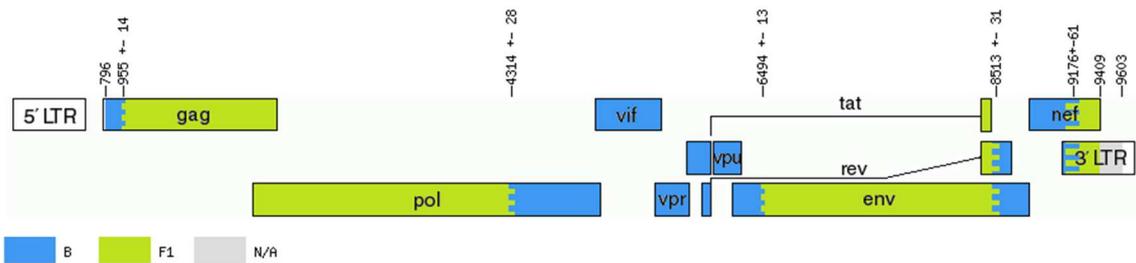
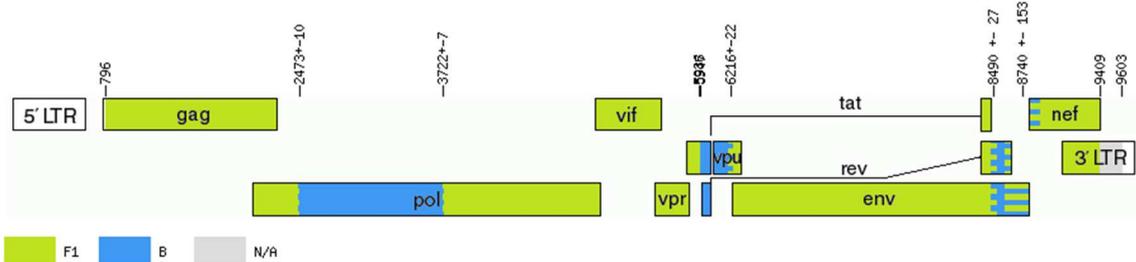


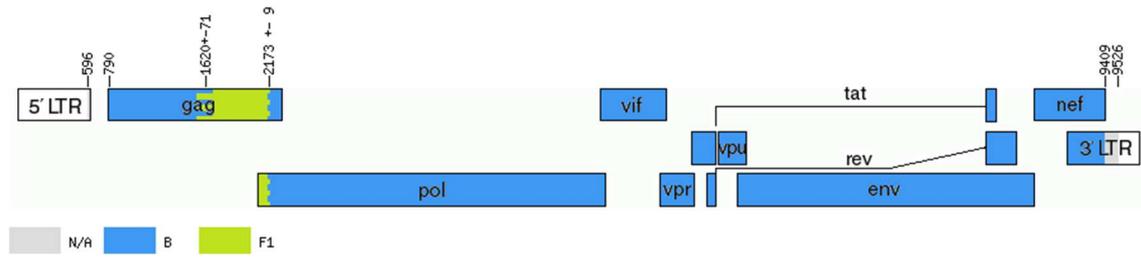
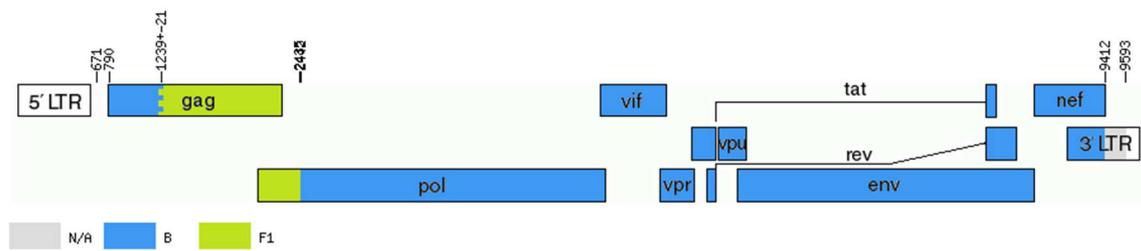
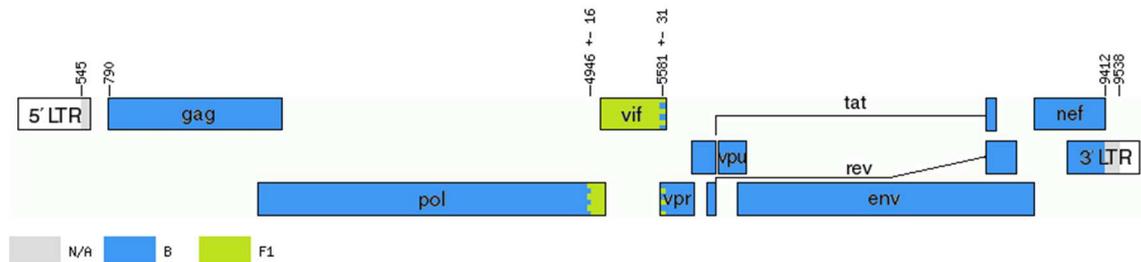
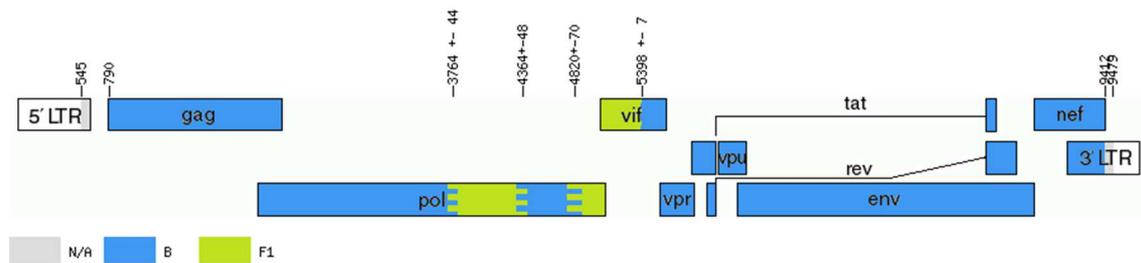
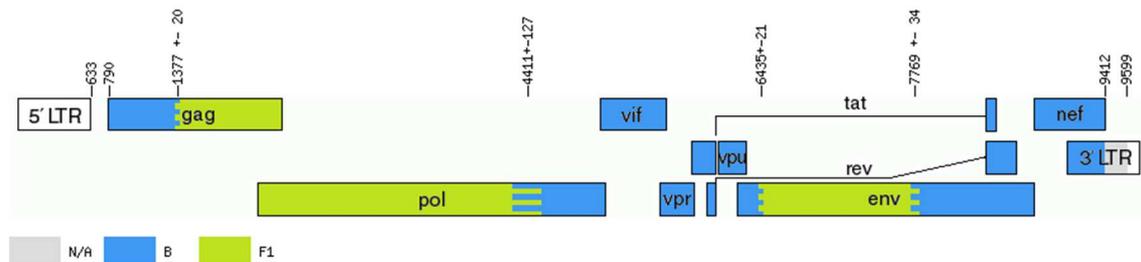
**JF804807.BR.2005.29****JF804808.BR.2005.BF1****JF804809.BR.2005.28****JF804810.BR.2005.BF1****JF804811.BR.2005.29**

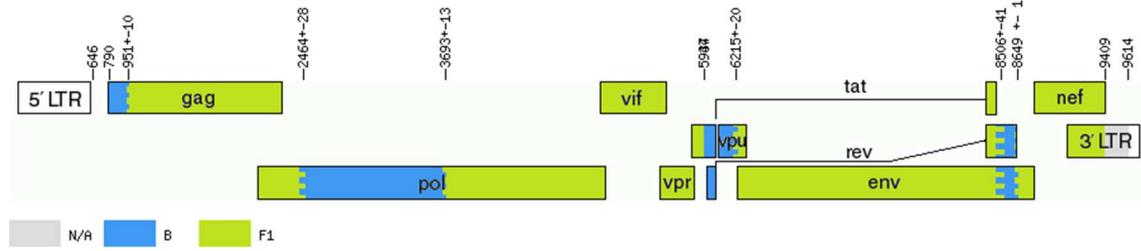
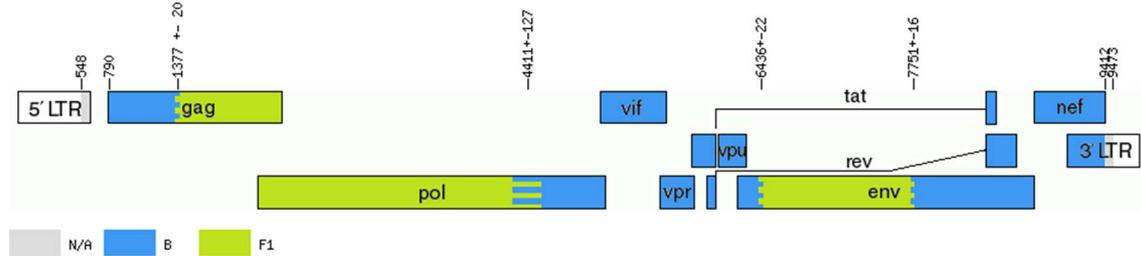
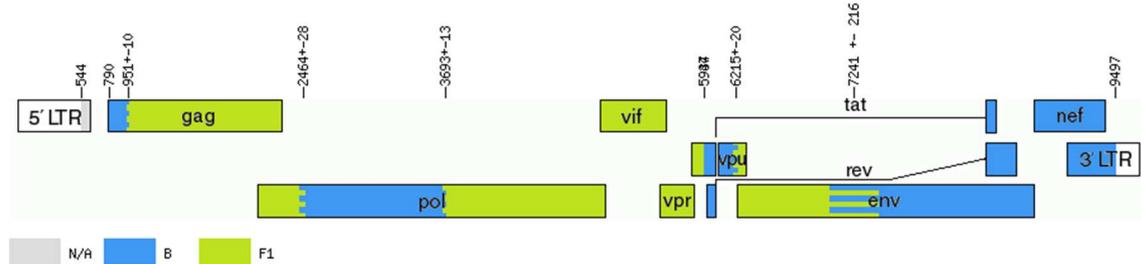
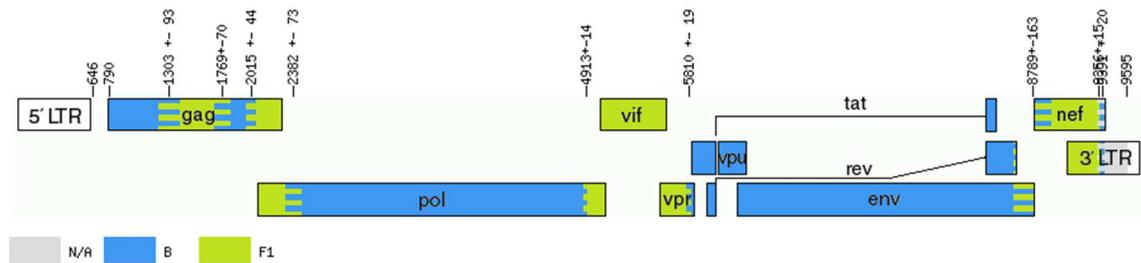
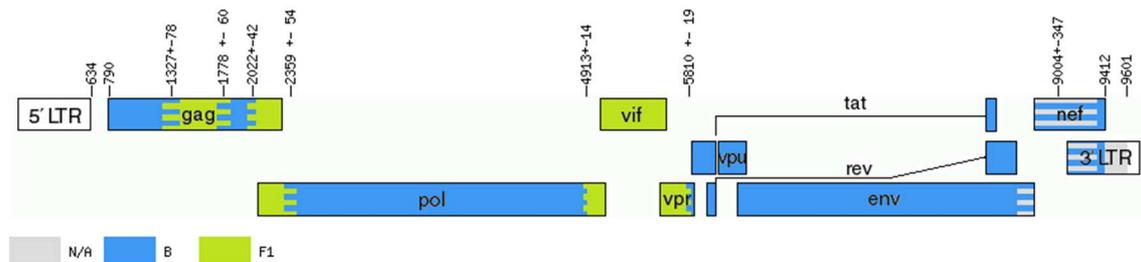
**JF804812.BR.2005.28****JF804813.BR.2005.BF1****JF804814.BR.2005.28****JN235952.UR.2000.BF****JN235953.UR.2001.BF**

**JN235954.UR.2001.BF****JN235955.UR.2001.BF****JN235956.UR.2001.BF****JN235960.UR.1999.BF****JN235962.UR.1999.38**

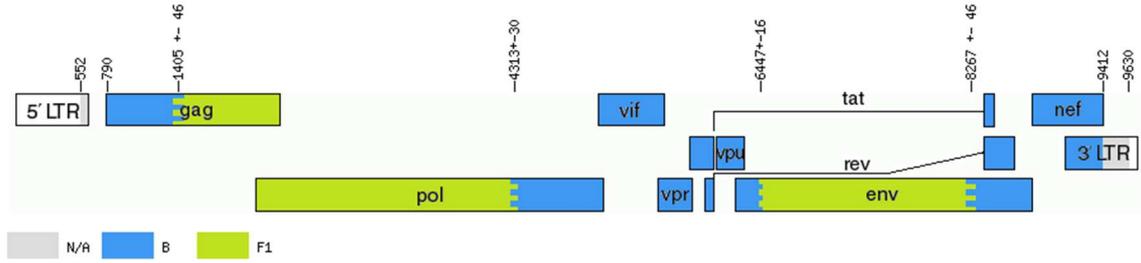
**JN235963.UR.1999.BF****JN235964.UR.1999.BF****JN251897.PR.2002.BF1****JN251898.PR.2002.BF1****JN251899.PR.2002.BF1**

**JN251900.PR.2002.BF1****JN251902.PR.2002.BF1****JN251903.PR.2002.BF1****JN251904.PR.2003.BF****JN251905.PR.2003.BF1**

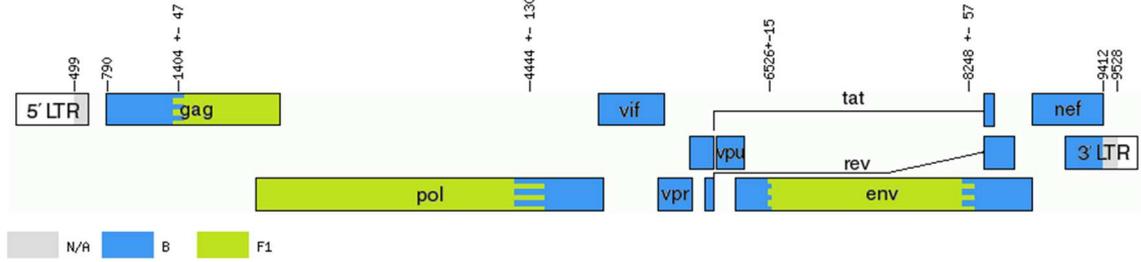
**JN692437.BR.2002.BF1****JN692448.BR.2003.BF1****JN692449.BR.2003.BF1****JN692456.BR.2004.BF1****JX140660.ES.2009.BF1**

**JX140661.ES.2009.BF1****KC113006.ES.2009.BF1****KC113008.ES.2009.BF1****KC473839.ES.2008.47****KC473840.ES.2008.47**

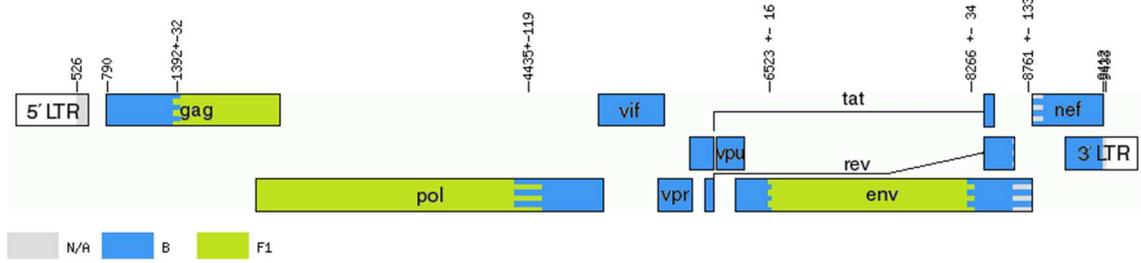
**KJ671533.BR.2010.72**



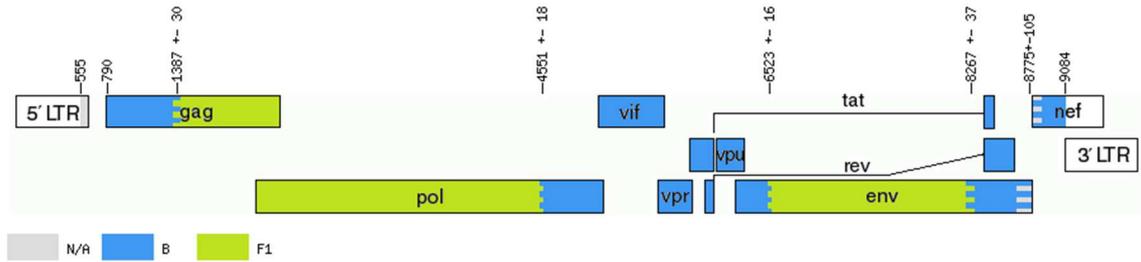
**KJ671534.BR.2010.72**



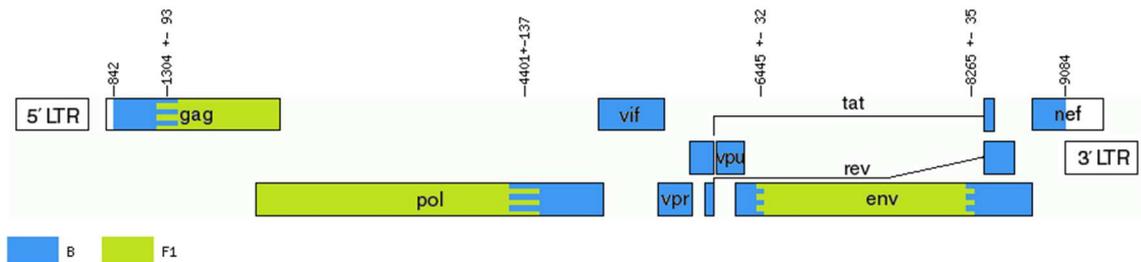
**KJ671535.BR.2010.72**

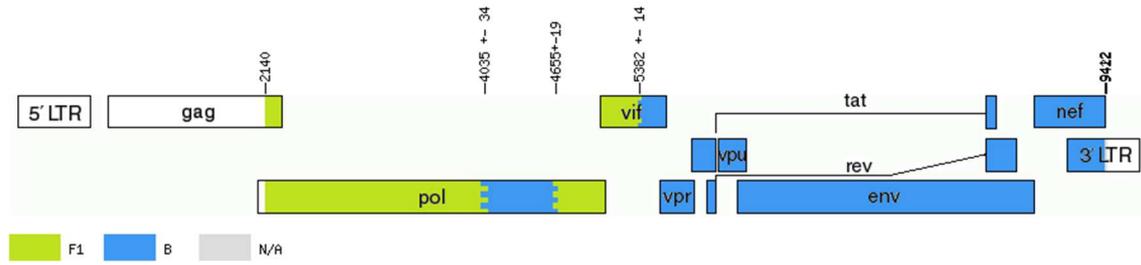
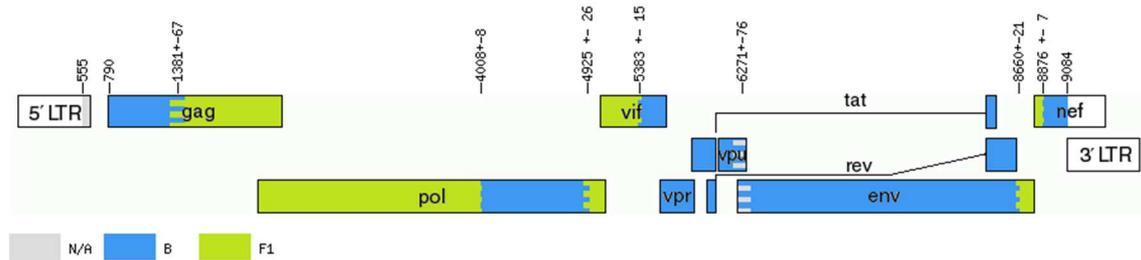
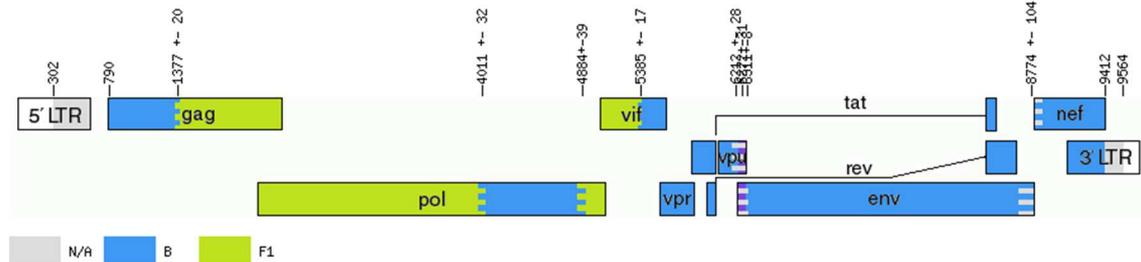
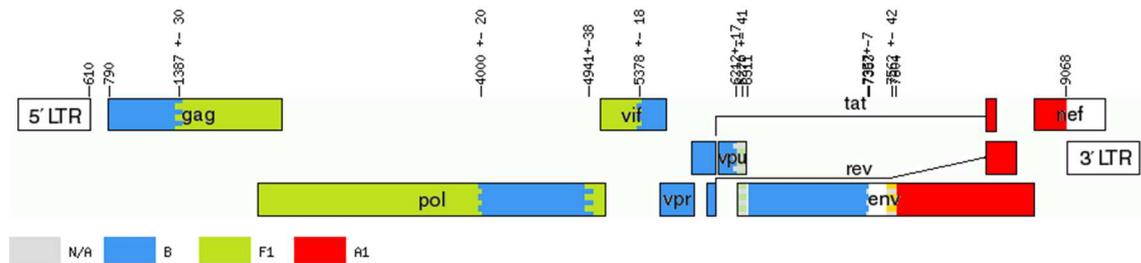
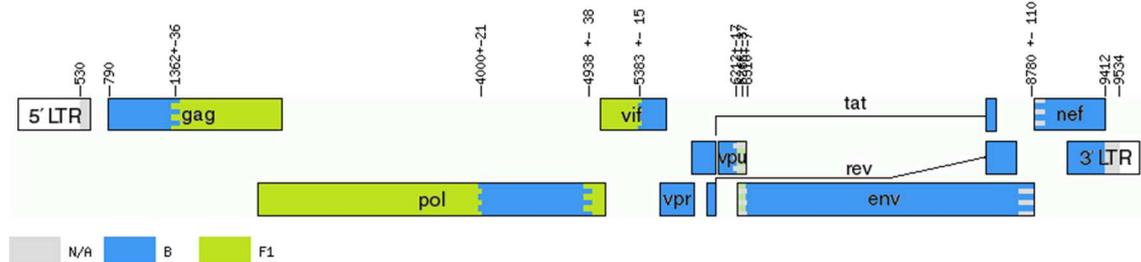


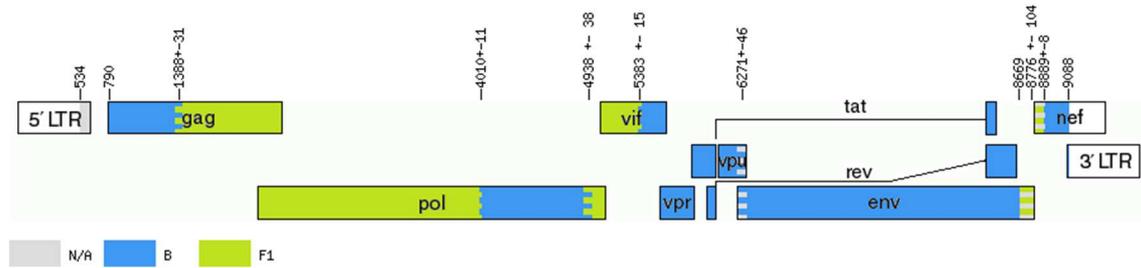
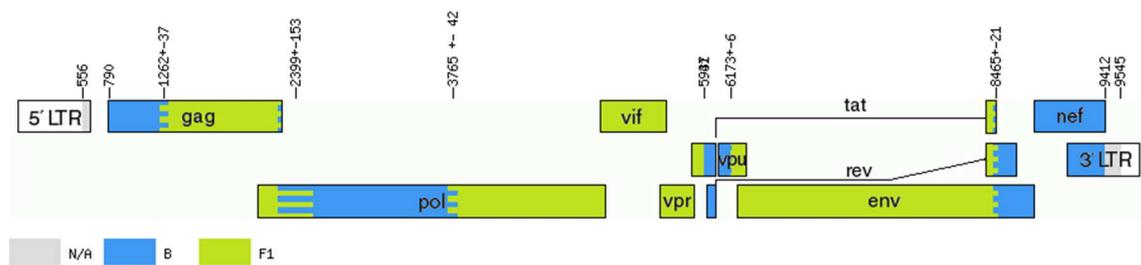
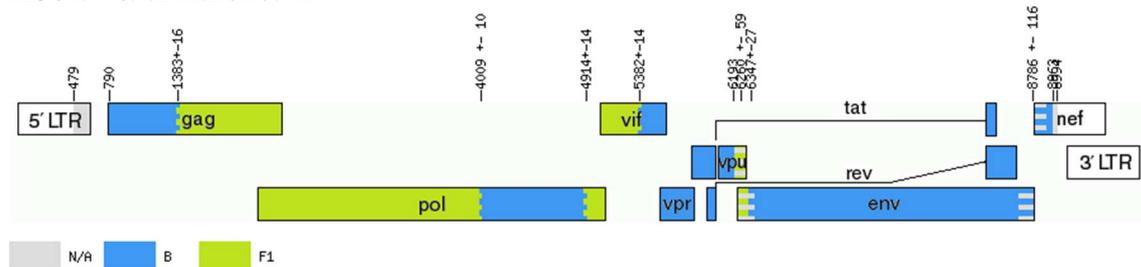
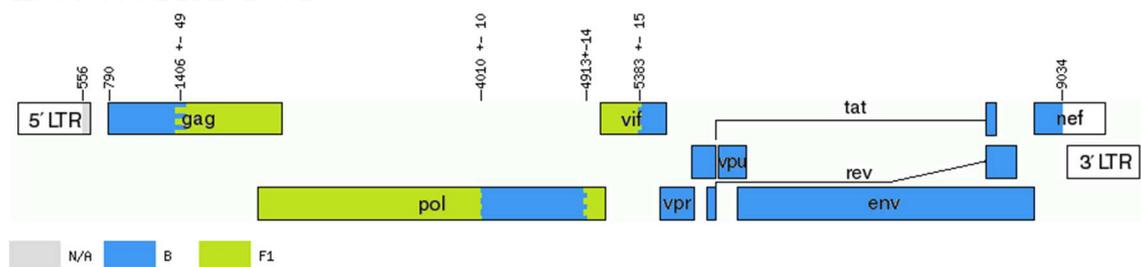
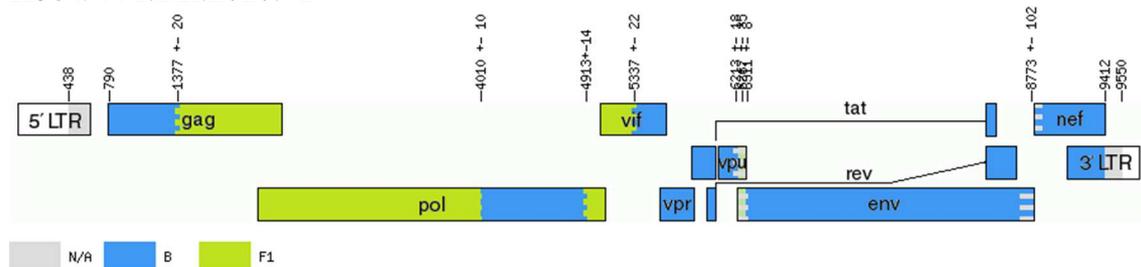
**KJ671536.BR.2010.72**

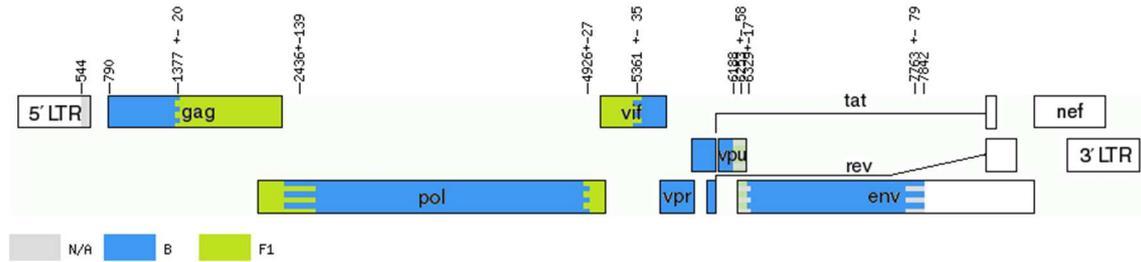
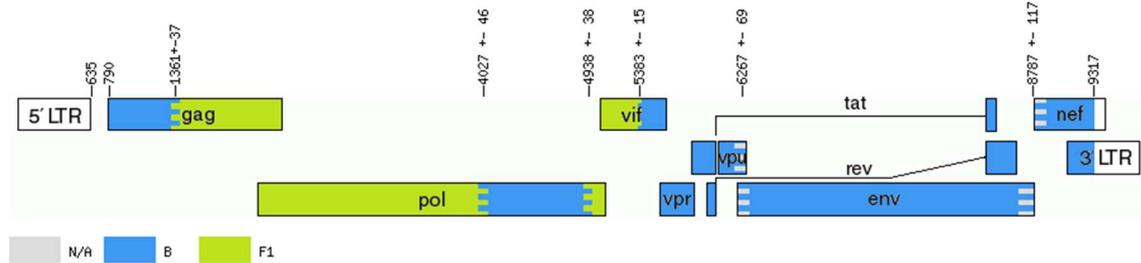
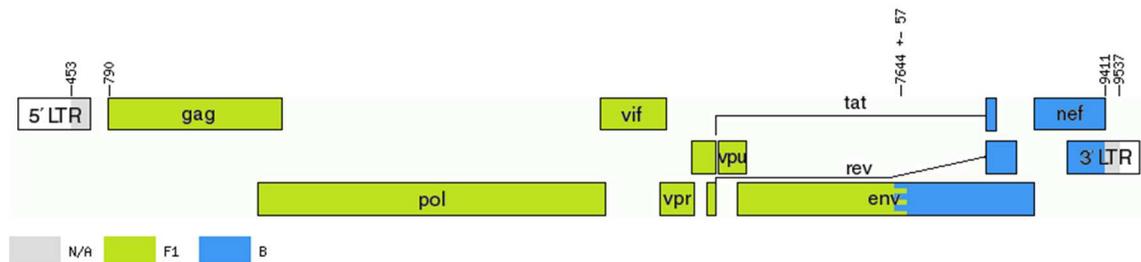
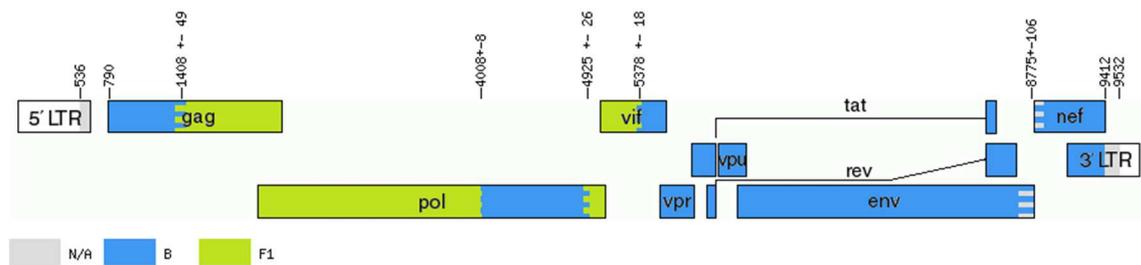
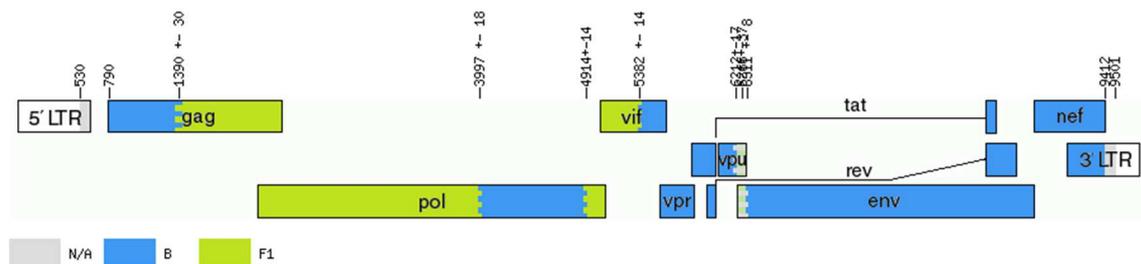


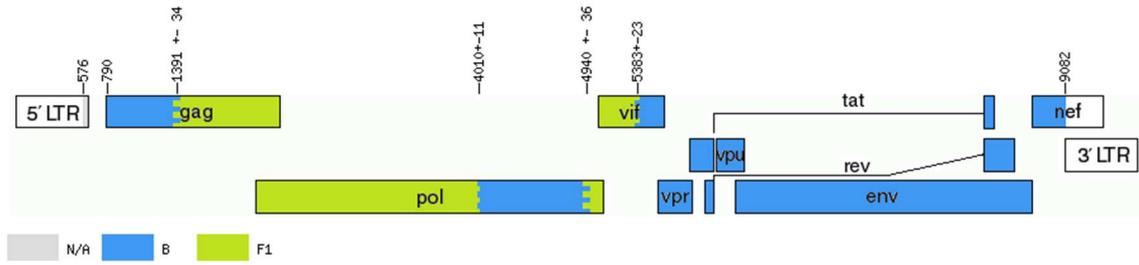
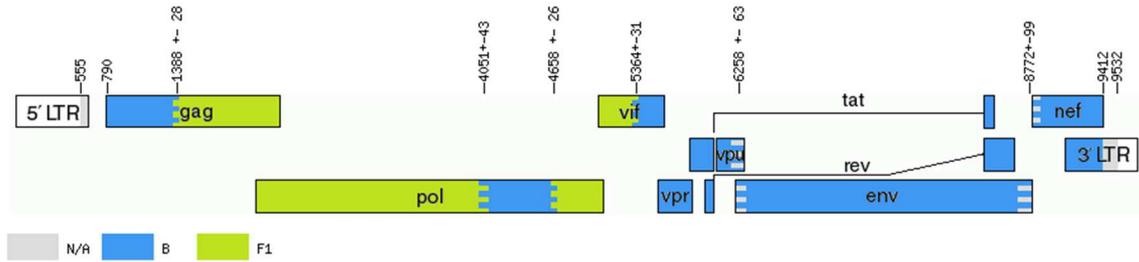
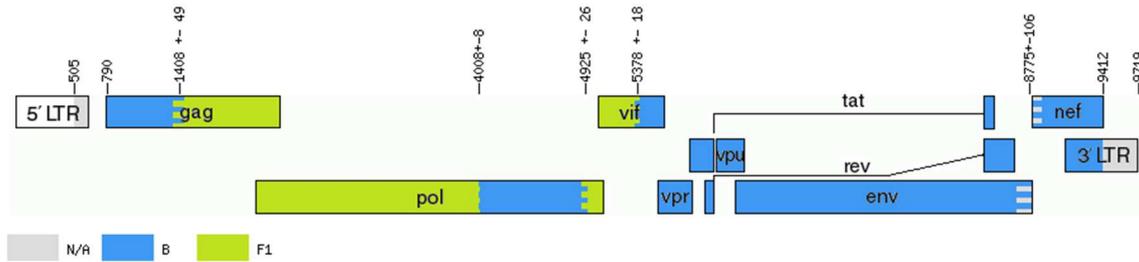
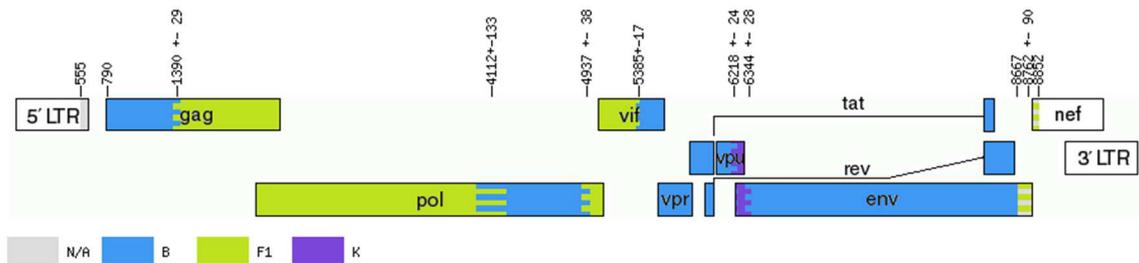
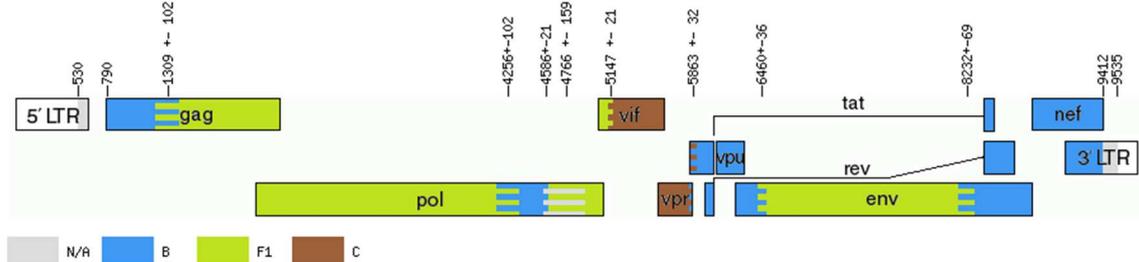
**KJ671537.BR.2010.72**

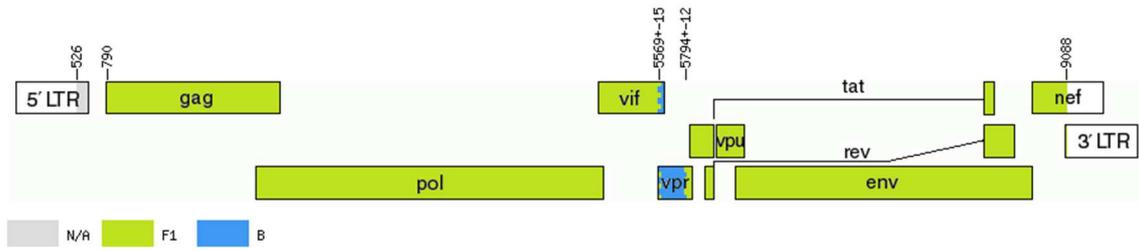
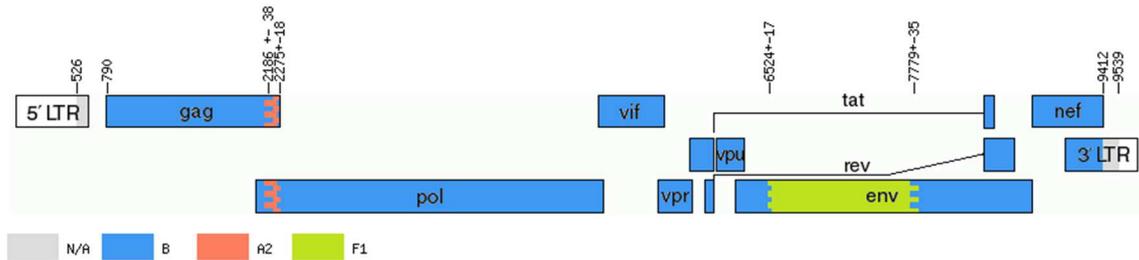
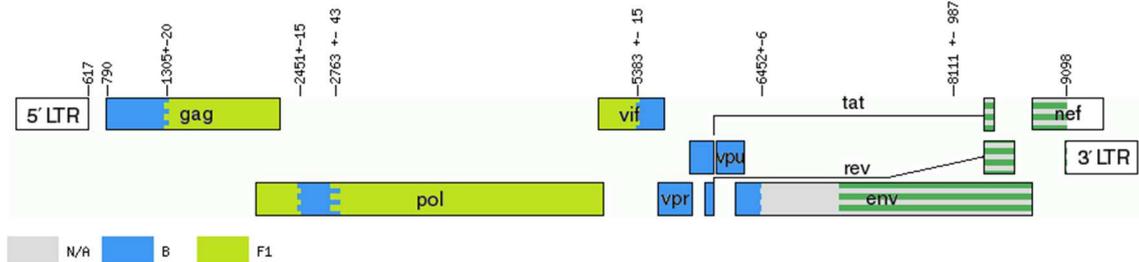
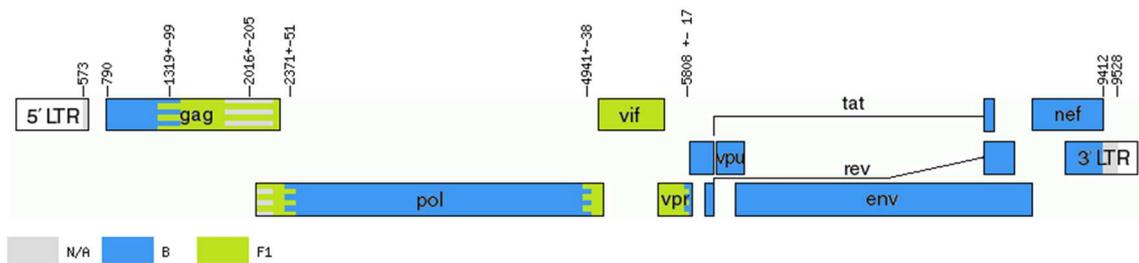
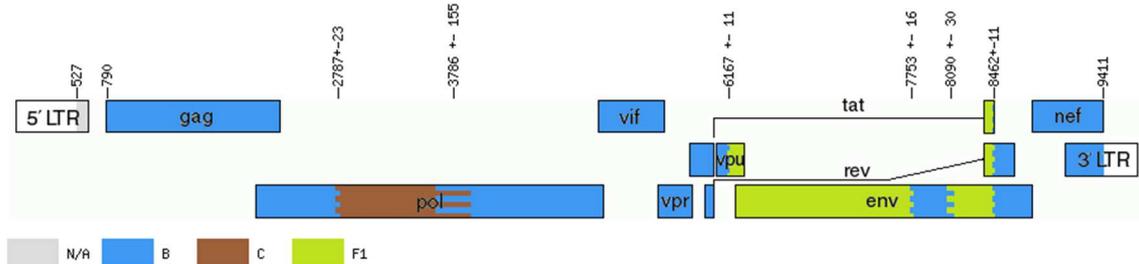


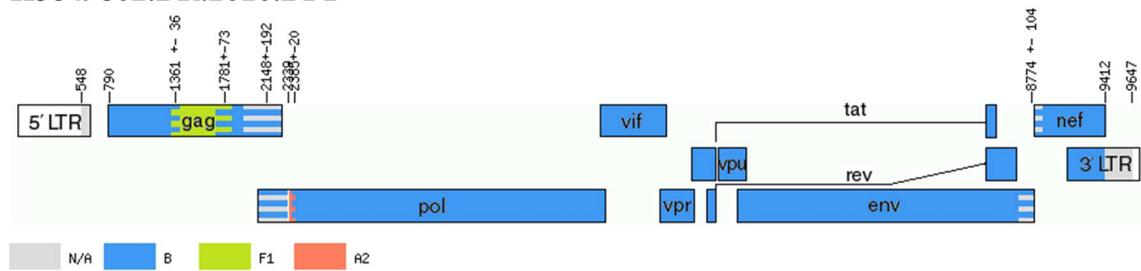
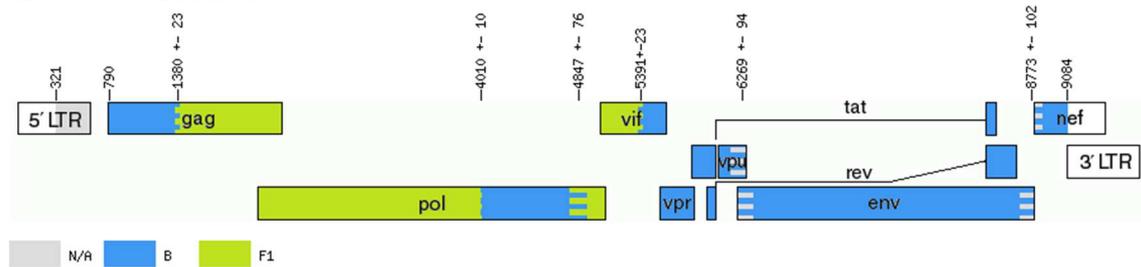
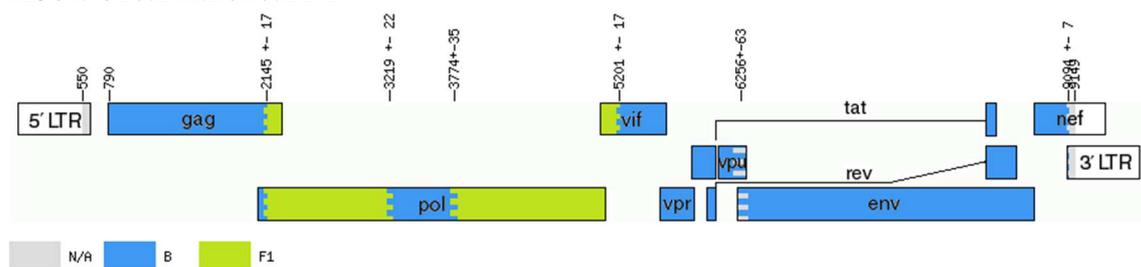
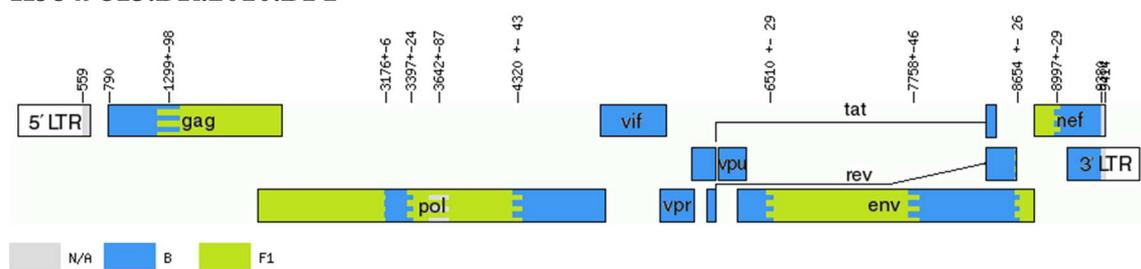
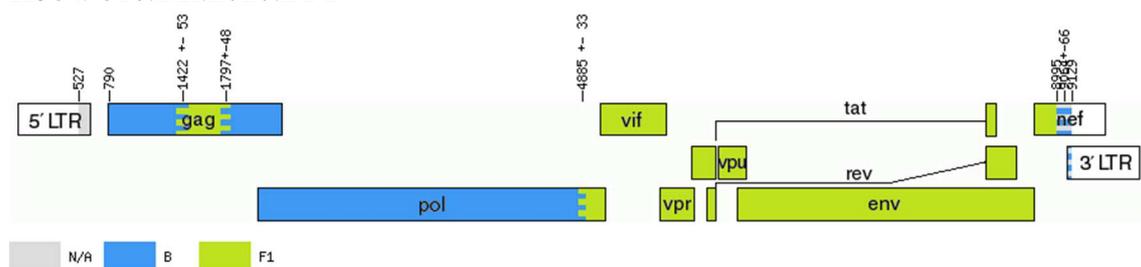
**KJ849757.BR.2010.BF1****KJ849758.BR.2010.70****KJ849759.BR.2010.71****KJ849760.BR.2010.71****KJ849761.BR.2010.70**

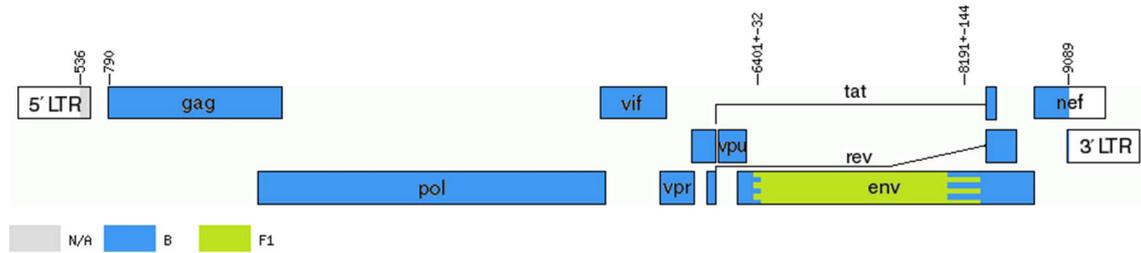
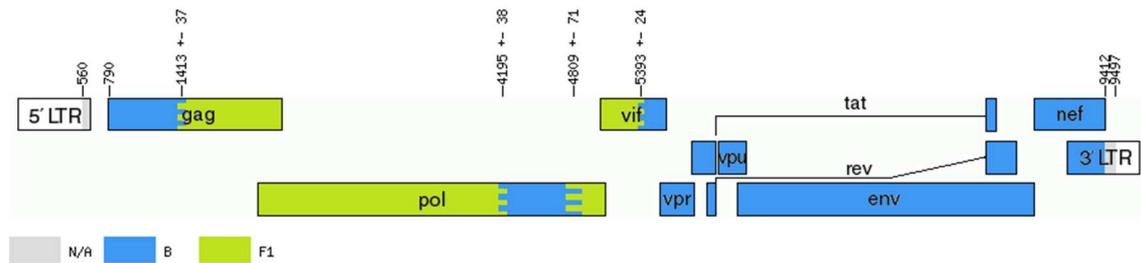
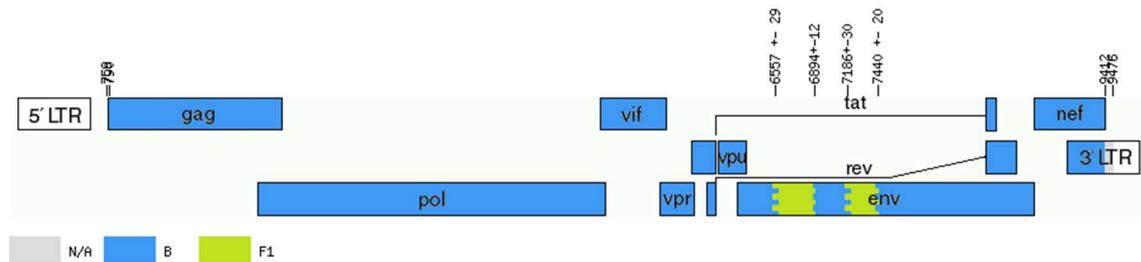
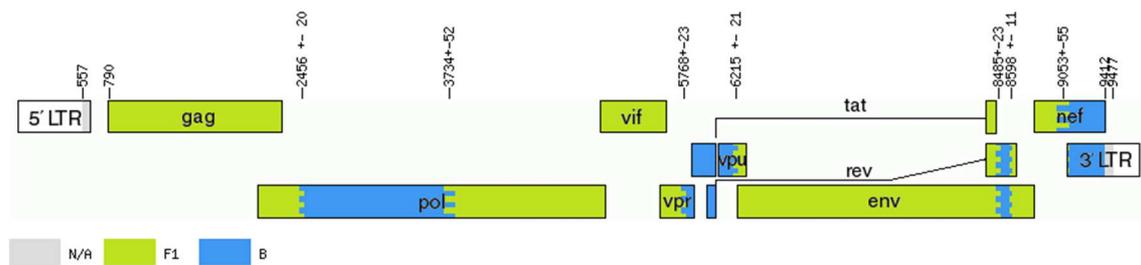
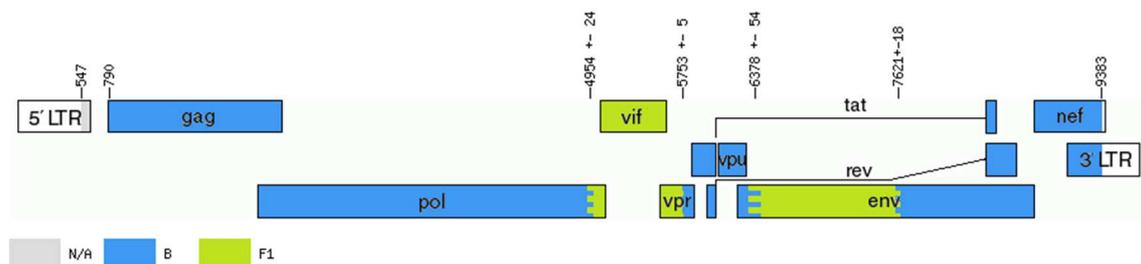
**KJ849762.BR.2010.70****KJ849768.BR.2010.BF1****KJ849769.BR.2010.71****KJ849770.BR.2010.71****KJ849771.BR.2010.71**

**KJ849772.BR.2010.BF1****KJ849773.BR.2010.71****KJ849774.BR.2010.BF1****KJ849775.BR.2010.71****KJ849776.BR.2010.71**

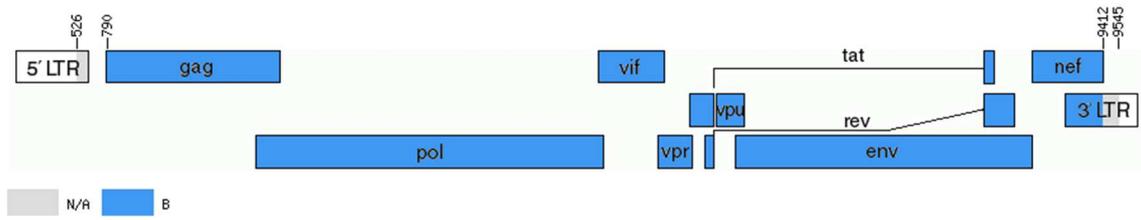
**KJ849777.BR.2010.71****KJ849778.BR.2010.71****KJ849779.BR.2010.71****KJ849783.BR.2010.70****KJ849787.BR.2010.BCF1**

**KJ849791.BR.2010.BF1****KJ849794.BR.2010.BF1****KJ849797.BR.2010.BF1****KJ849798.BR.2010.47****KJ849800.BR.2010.BCF1**

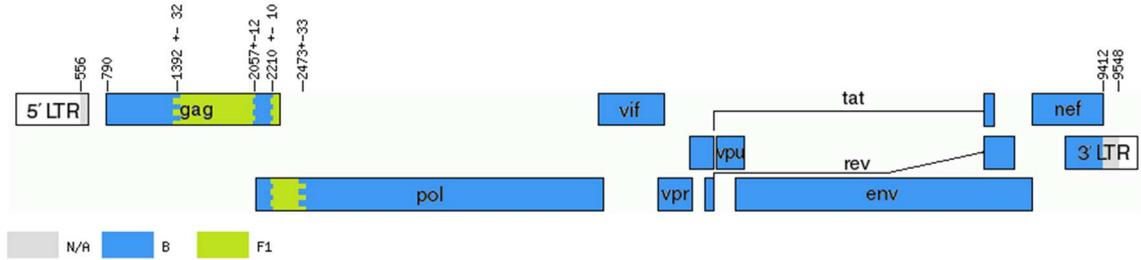
**KJ849802.BR.2010.BF1****KJ849809.BR.2010.70****KJ849810.BR.2010.BF1****KJ849813.BR.2010.BF1****KJ849816.BR.2010.BF1**

**KJ849826.BR.2010.BF1****KT276259.ES.2014.BF1****KT276269.ES.2014.BF1****KT276271.ES.2014.BF1****KT427649.BR.2010.BF1**

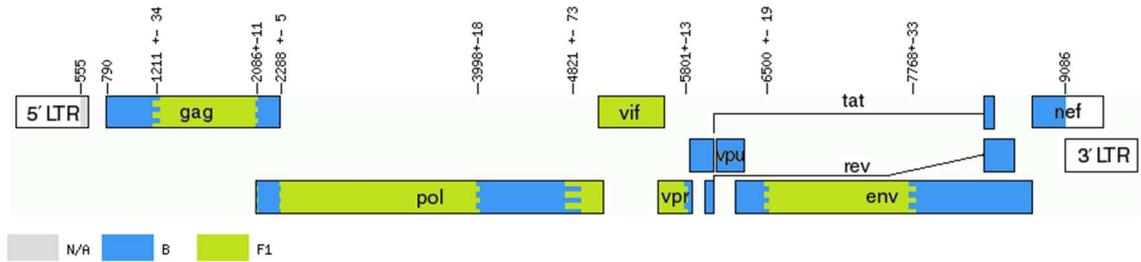
**KT427650.BR.2010.BF1**



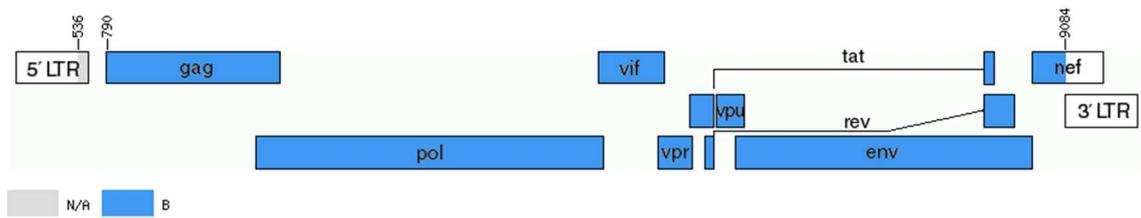
**KT427651.BR.2010.BF1**



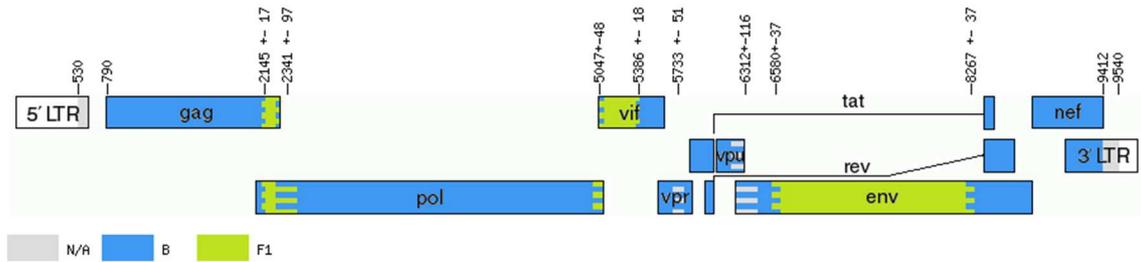
**KT427652.BR.2010.BF1**

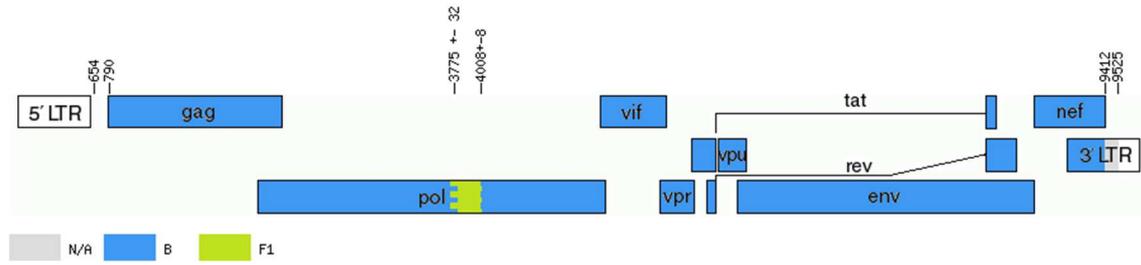
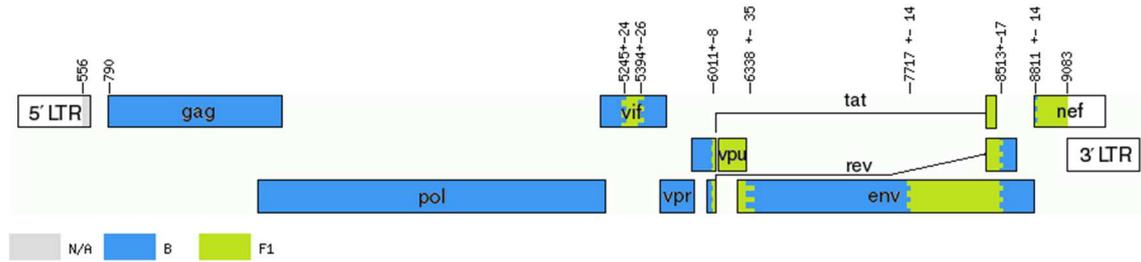
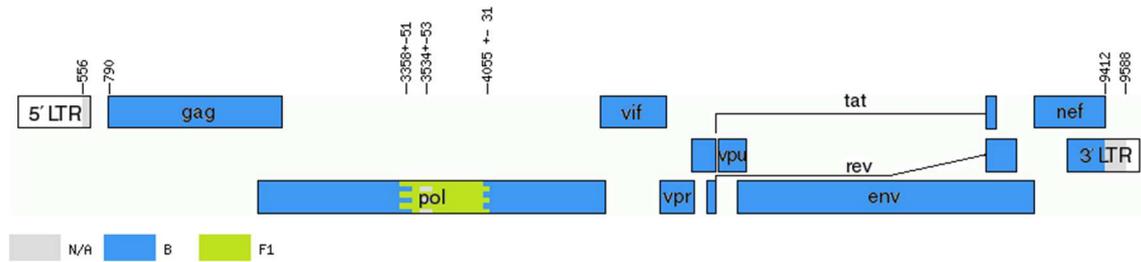
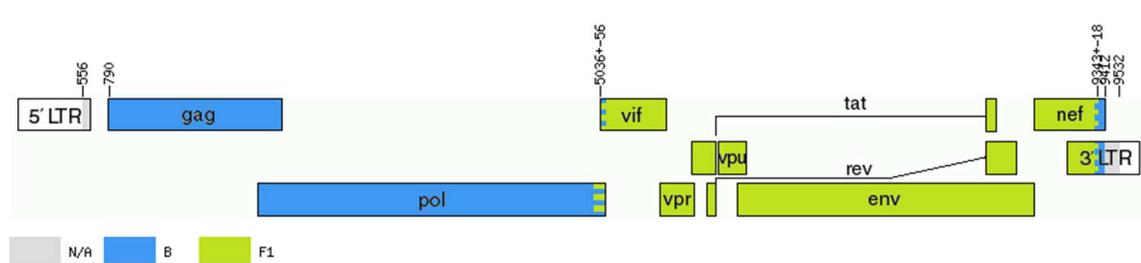
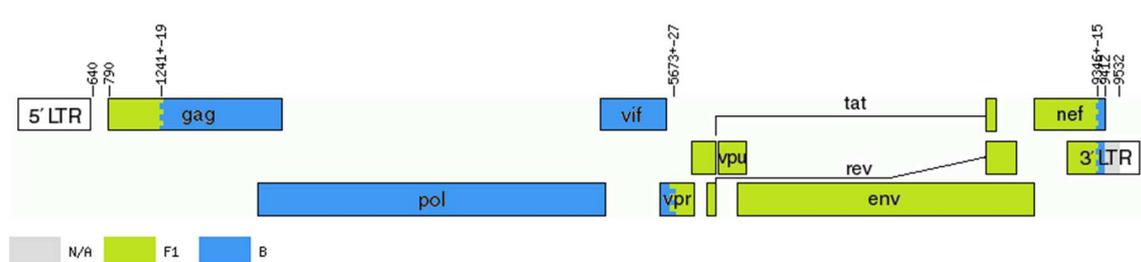


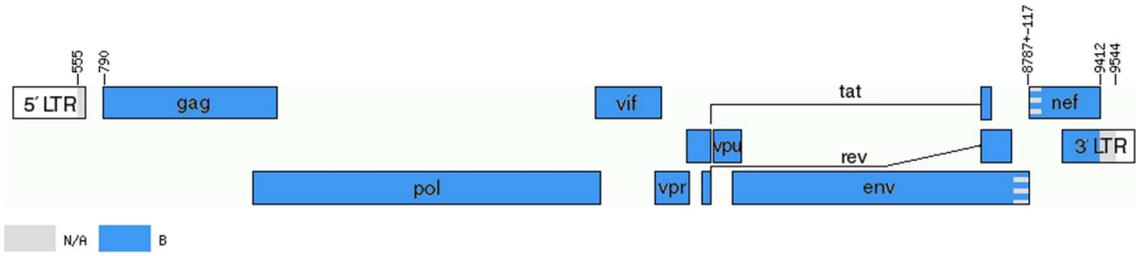
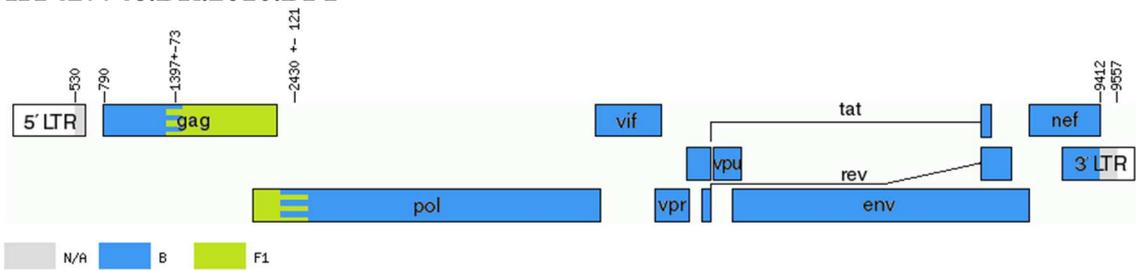
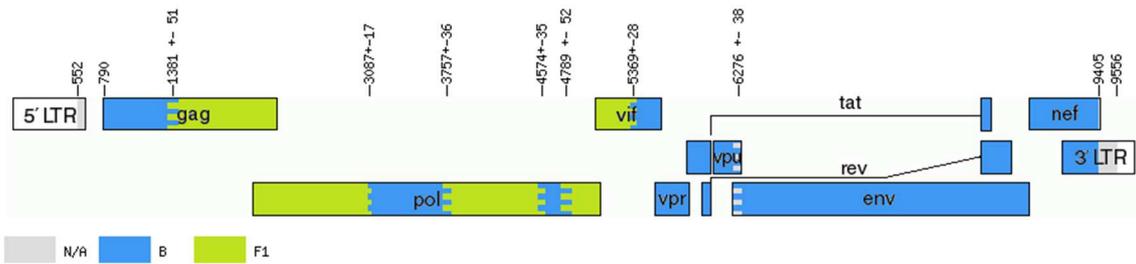
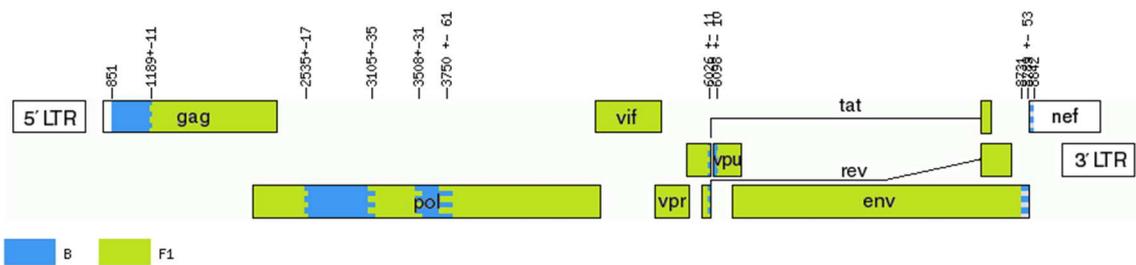
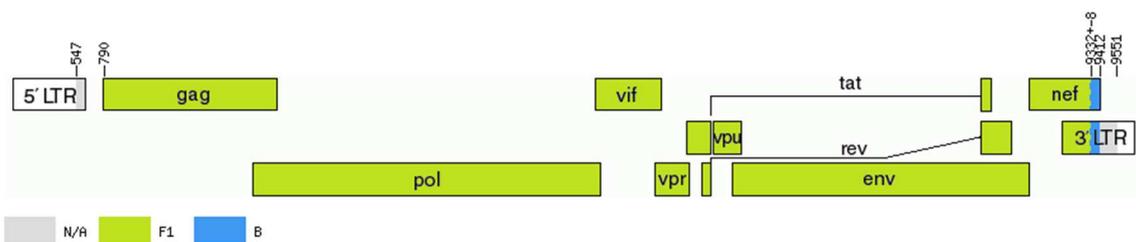
**KT427668.BR.2010.BF1**



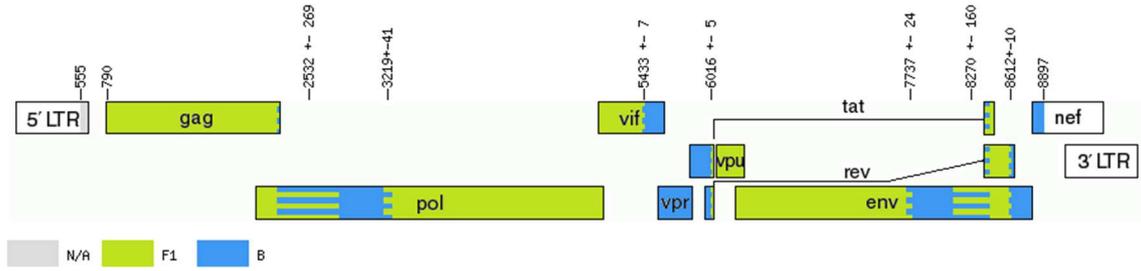
**KT427672.BR.2010.BF1**



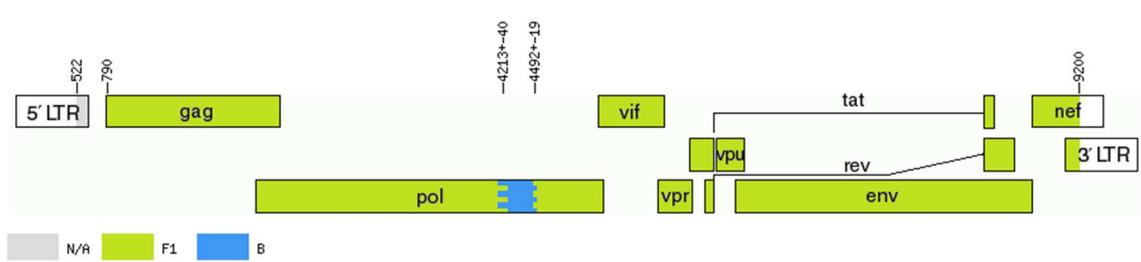
**KT427679.BR.2010.BF1****KT427690.BR.2010.BF1****KT427698.BR.2010.BF1****KT427705.BR.2010.BF1****KT427710.BR.2010.BF1**

**KT427722.BR.2010.BF1****KT427746.BR.2010.BF1****KT427758.BR.2010.BF1****KT427759.BR.2010.BF1****KT427774.BR.2010.BF1**

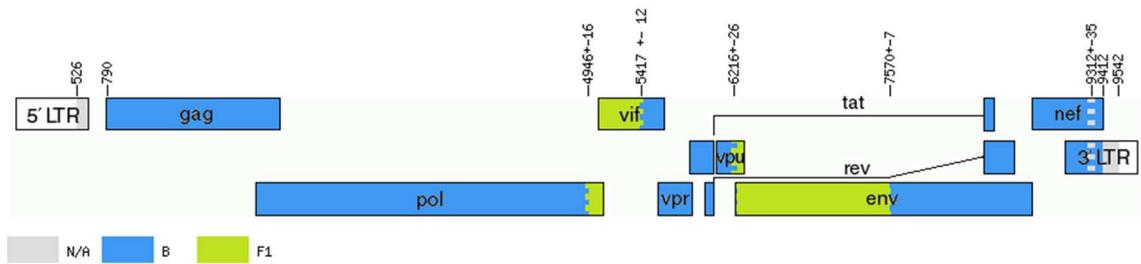
**KT427776.BR.2010.BF1**



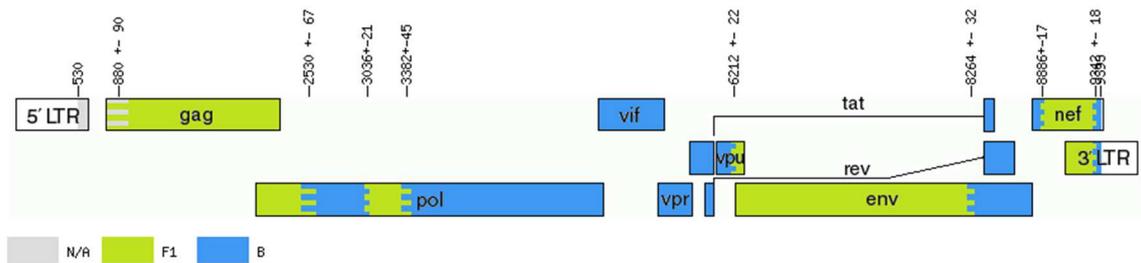
**KT427779.BR.2010.BF1**



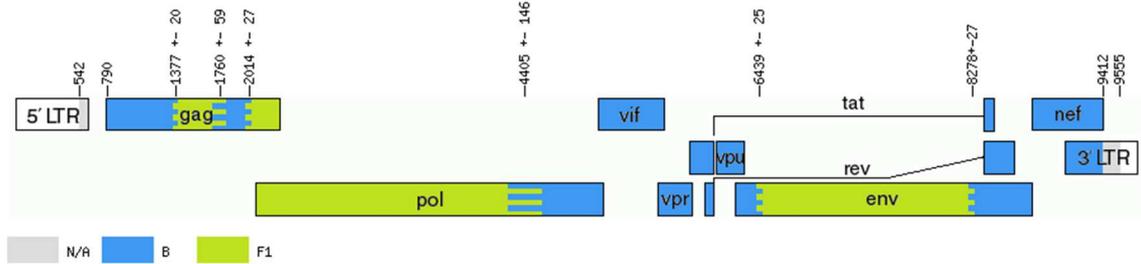
**KT427780.BR.2010.BF1**

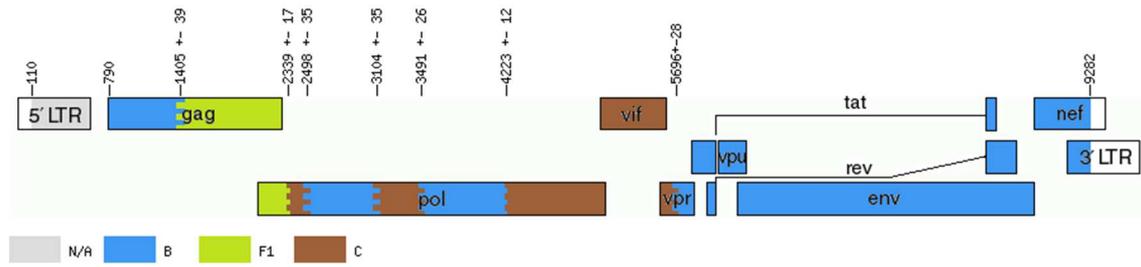
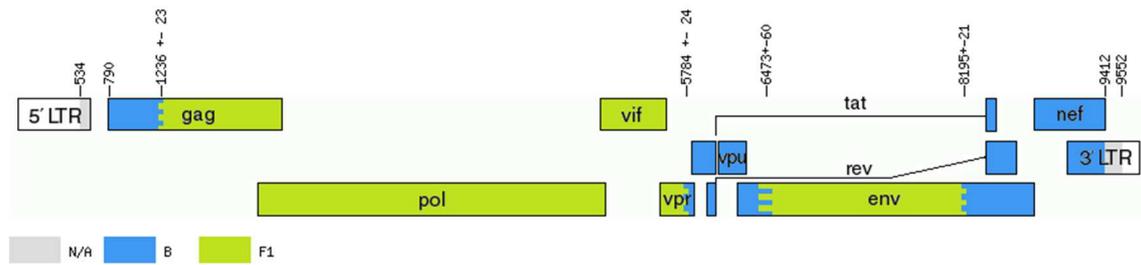
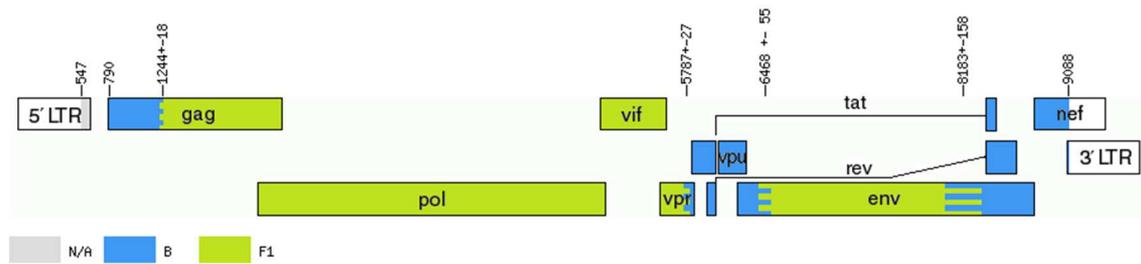
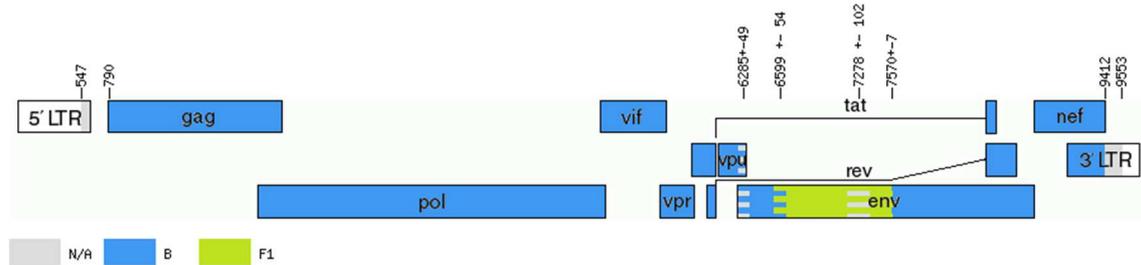
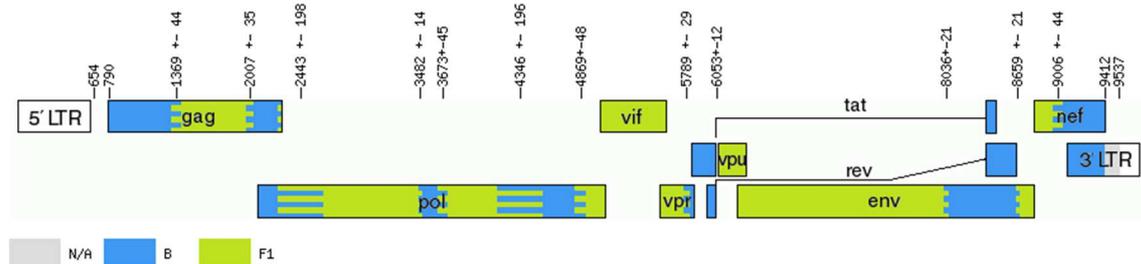


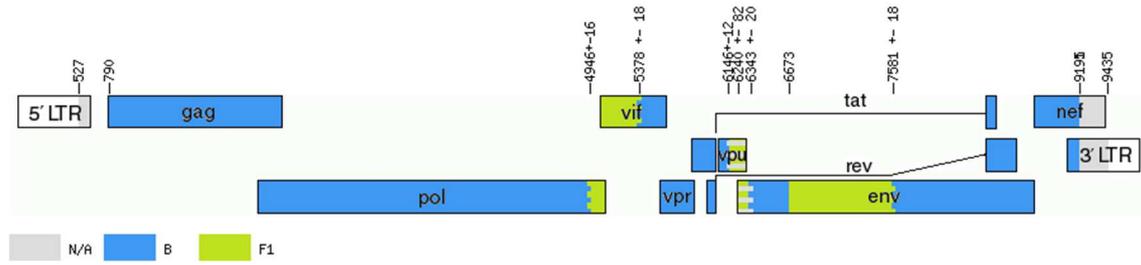
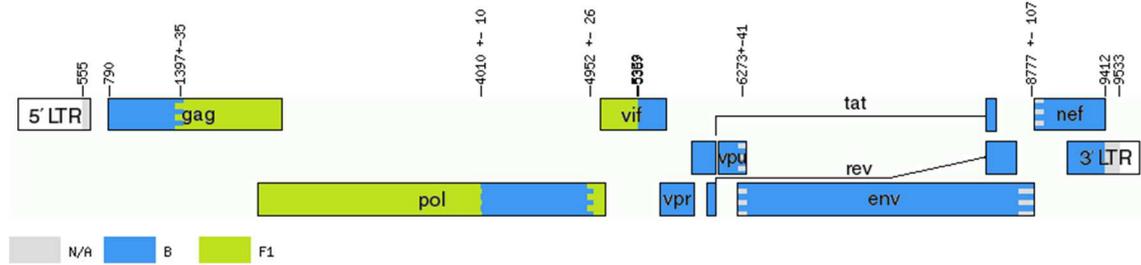
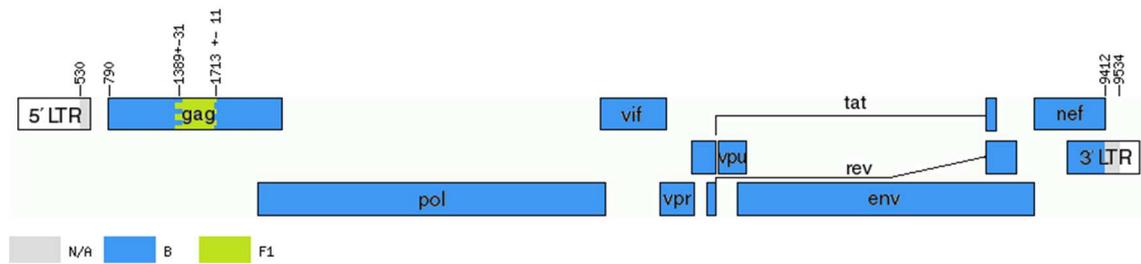
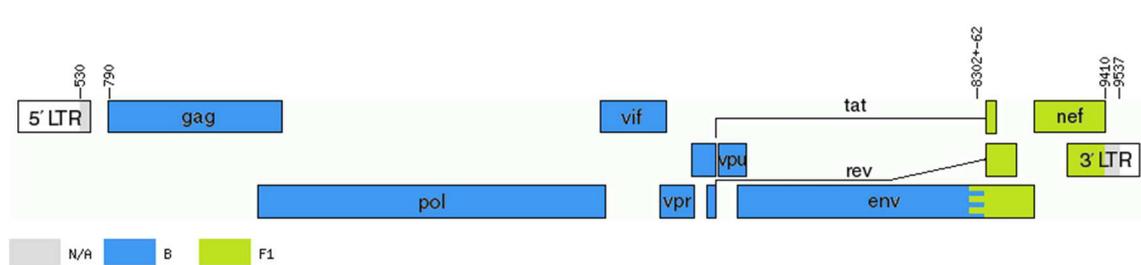
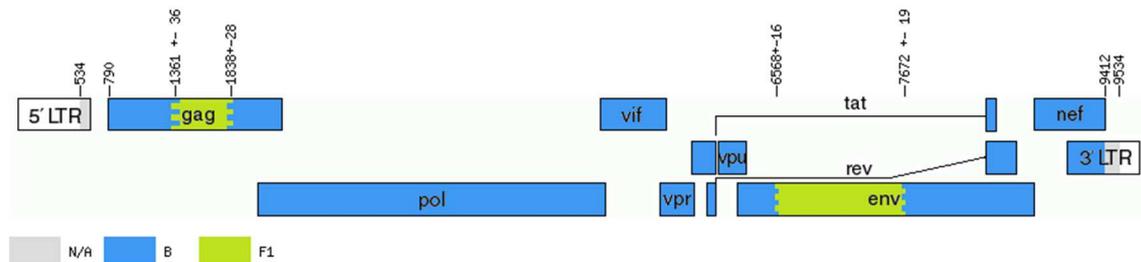
**KT427782.BR.2010.BF1**

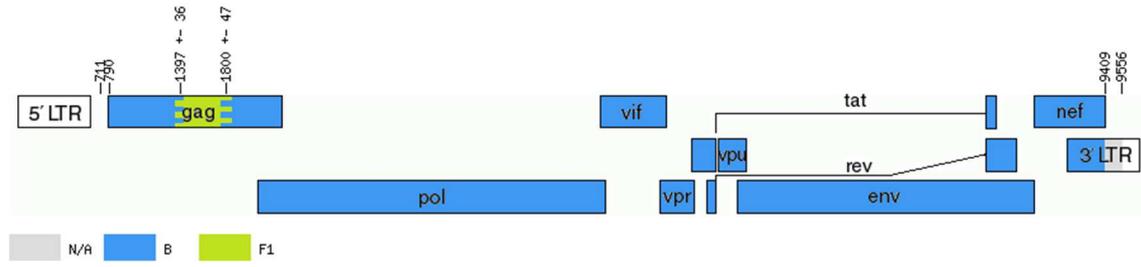
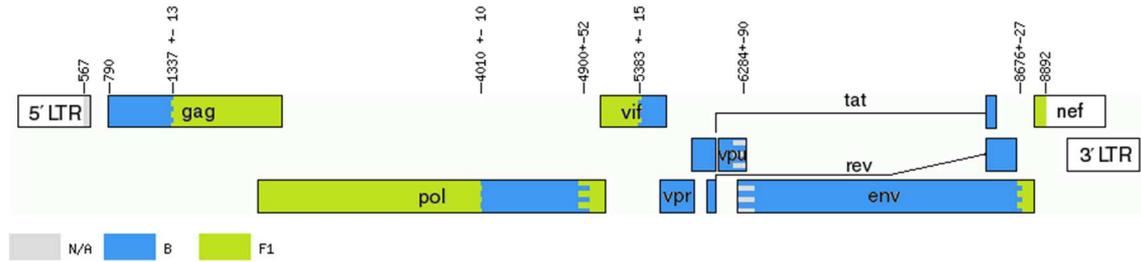


**KT427789.BR.2010.BF1**



**KT427793.BR.2010.BCF1****KT427795.BR.2010.BF1****KT427798.BR.2010.BF1****KT427801.BR.2010.BF1****KT427807.BR.2010.BF1**

**KT427809.BR.2010.BF1****KT427816.BR.2010.71****KT427820.BR.2010.BF1****KT427823.BR.2010.BF1****KT427825.BR.2010.BF1**

**KT427828.BR.2010.BF1****KT427871.BR.2010.BF1**

**APÊNDICE D** – Assinaturas nucleotídicas conservadas exclusivas de cada CRF. N: número de sequências de cada CRF.

<b>CRF</b>	<b>N</b>	<b>Posições com Assinaturas nucleotídicas conservadas exclusivas (n)</b>	<b>Posições com AMEC/base conservada</b>
CRF12	6	10	3296A/4355G/4406G/5118G/6023A/7352C/7466T/8777G/8916G/8946C
CRF17	7	4	2843T/3032G/3800T/9108A
CRF28	6	27	145T/5244T/5875G/6321C/6405G/6672A/6705G/6715G/6765G/6857A/6920A/7047G/7241A/7501G/7651G/7973A/8208C/8471A/8826G/8859aA/8859bC/8859cC/8859dA/8860G/8880T/9243T/9402T
CRF29	7	11	1714T/1715G/1858A/7035G/7053T/7273G/7541C/8155C/8369A/8719T/8842G
CRF38	5	15	1047C/1090C/1132A/1221C/2903A/3856A/5799G/5838T/6021T/6061fA/6061gT/6504G/7624T/7771C/9270G
CRF39	3	81	878G/1163A/1167T/1428T/1434G/1443A/1458T/1459G/1476A/1482A/1494C/1500A/1548T/1623T/1636C/1659G/1684C/1691A/1698T/1707C/1717T/1753C/1809A/1851G/1953G/1962G/1986G/2043G/2133T/2179A/2187G/2195C/2199A/2202T/2206T/2207C/2230A/2234A/2519C/2952G/3080A/3099G/3147G/3192T/3392G/3464G/3470A/3678T/3679T/3959G/4200C/4211A/4217T/4421C/4505G/4599A/4929C/4952T/4979A/5283A/5732G/5829G/5899T/5920C/6041G/6089A/6383C/6530G/6645G/6651C/6758G/6789G/7227G/7618A/7820G/7887C/8090C/8105A/8396A/8607C/9174T
CRF40	4	74	864G/909A/910T/1146C/1151C/1169G/1170C/1171A/1287G/1326T/1362A/1432C/1569G/2064T/2067T/2073G/2311A/3066G/3174T/3482G/3635C/3722G/4148A/4373G/4431C/4607C/4842T/4925T/5289A/5415G/5436T/5504A/5576G/5679A/5721A/6070C/6081C/6093G/6140C/6151C/6226A/6314C/6622C/6623C/6624T/6625T/6749C/6773T/6965C/7480A/75

			61A/7729G/7764A/7765C/8080T/8120G/8143G/8504A/8606G/8677T/8699A/8821C/8824G/8827C/8828C/8829C/8863G/8894C/8910G/9123A/9129C/9138T/9141T/9384C
CRF42	21	99	807G/954G/960C/1039A/1040C/1315G/1395G/1635C/1695G/1699T/1909G/1946C/2046G/2235T/2302A/2357A/2360G/2374G/2414T/2439A/2529C/2549C/2847C/2953C/3148T/3189C/3344C/3363G/3365G/3368A/3380A/3473G/3590G/3613G/3692C/3734G/3737A/3757T/3839G/4005C/4043T/4061G/4233C/4271G/4311A/4344T/4603T/4742C/4745C/5015G/5128T/5156A/5228G/5240A/5407A/5461C/5681T/5684G/5760T/5789G/5805T/5811C/6000A/6001C/6009C/6017G/6031G/6127A/6235T/6306A/6310G/6338T/6378T/6410c/7112C/7115G/7187C/7229G/7231A/7279T/7325A/7335G/8076A/8095G/8123G/8142C/8207A/8246T/8259T/8395A/8462A/8464C/8472G/8525T/8663T/8664c/8738G/8750T/8758C
CRF44	2	129	849A/880A/885G/889C/892A/928G/978A/1016G/1023G/1049T/1074G/1141A/1142A/1144A/1147C/1148A/1150C/1155G/1157C/1162G/1228C/1320G/1329T/1532G/1782A/1789C/1872G/2040G/2052T/2236A/2287A/2555C/2669G/2684G/2756C/2822G/2906A/3077G/3377G/3393T/3395A/3414T/3626G/3989C/4096G/4128G/4220A/4320A/4323A/4630A/4661C/4716A/4861G/5039C/5160C/5191G/5368C/5419T/5447C/5488T/5500G/5621A/5641G/5682C/5747A/5915C/5936T/5937G/5947G/5950G/6055G/6061dT/6068G/6083A/6117T/6158C/6160G/6293A/6478A/6479T/6628C/6668A/6717G/6733T/6757G/6775aC/6784G/6785C/6829A/6943C/7065A/7168A/7214C/7254A/7289C/7294C/7390G/7441C/7446T/7479C/7553A/7554T/7615C/7622A/7634A/7882G/7883A/7894T/8075G/8111A/8384T/8385A/8530G/8586G/8697T/8788A/8866C/8890A/8891A/8900A/9044C/9049G/9060T/9094T/9213G/9367C/9369G/9389G/9406C
CRF46	7	17	858G/1025T/1323G/1903C/2303G/2480G/2567C/2608G/3143C/3293C/3428G/3438G/3890C/8608T/8623C/9340C/9341A

CRF47	5	177	828G/861A/1065C/1201G/1272A/1390C/1509C/1646A/1683T/1725T/1765T/1807C/1808A/1842A/1881A/1906T/1910A/1913T/1915A/1937A/2034G/2307A/2422G/2435A/2457A/2528A/2612C/2630G/2780T/2876A/2910T/2913G/3029T/3071G/3086G/3227C/3379G/3389C/3391G/3578G/3646G/3650G/3701G/3704A/3821G/3852A/3866T/3927A/3929T/3931A/3968C/3997T/4097A/4130C/4298C/4334G/4514G/4530C/4604A/4691A/4697G/4830G/5127A/5132C/5157C/5226A/5226aA/5226cT/5265G/5318A/5506A/5523G/5594A/5639A/5788A/5917A/6048G/6113T/6183A/6253G/6273G/6418C/6476G/6593A/6599T/6608C/6626T/6676G/6686G/6713C/6714C/6718C/6719G/6725G/6727G/6728A/6729A/6732A/6770G/6772A/6774T/6777G/6846A/6918C/7085G/7098A/7101A/7102G/7122G/7123G/7169G/7171T/7197A/7199T/7228A/7244C/7252G/7261C/7264G/7303T/7311C/7312A/7338G/7348C/7355C/7386A/7418G/7419A/7421T/7422G/7423A/7425A/7427A/7429A/7431A/7432T/7439A/7440G/7443T/7444C/7447A/7463A/7543G/7544T/7548T/7555A/7570C/7580G/7581A/7588T/7710A/7767T/7830G/8058A/8079T/8132T/8156C/8164C/8166C/8324A/8416G/8467C/8470C/8473G/8496A/8598C/8599T/8657C/8802A/8803A/8804A/8925C/8937C/9102T/9107G/9318G/9408G
CRF70	4	13	2372G/2837C/5591A/5687G/5839G/5911C/6716C/6884G/7049C/7056A/7256A/7763A/8213A
CRF71	12	3	2537T/4121G/6808G
CRF72	5	13	1605G/2735T/3272G/3624T/4889C/5178T/5232G/5669G/7005C/7895A/8360T/8609C/8945A



**Universidade Federal da Bahia  
Instituto de Ciências da Saúde**

**Programa Multicêntrico de Pós-Graduação em Bioquímica e Biologia Molecular  
Avenida Reitor Miguel Calmon s/n - Vale do Canela CEP: 40110-100  
Salvador, Bahia, Brasil  
Sociedade Brasileira de Bioquímica e Biologia Molecular**